ORIGINAL

Darrell Olson (State Bar No. 77633)
KNOBBE, MARTENS, OLSON & BEAR, LLP
2040 Main Street
Fourteenth Floor
Irvine, CA 92614
Phone: (949) 760-0404
Facsimile: (949) 760-9502

Joseph M. Reisman (State Bar No. 246922)
Ali S. Razai (State Bar No. 196122)
KNOBBE, MARTENS, OLSON & BEAR, LLP
550 West C Street
Suite 1200
San Diego, CA 92101
Phone: (619) 235-8550
Facsimile: (619) 235-0176

Attorneys for Plaintiff
LIGAND PHARMACEUTICALS INCORPORATED

IN THE UNITED STATES DISTRICT COURT

FOR THE SOUTHERN DISTRICT OF CALIFORNIA

'08 CV 401 BEN WMc

LIGAND PHARMACEUTICALS INCORPORATED, a Delaware corporation, Plaintiff, v. THE ROCKEFELLER UNIVERSITY, a New York not-for-profit corporation, Defendant.	Civil Action No. **COMPLAINT FOR DECLARATORY JUDGMENT**

I. NATURE OF THE ACTION

1. This is a civil action under the Declaratory Judgment Act, 28 U.S.C. § 2201, et seq., for declaration of rights between the parties under a License Agreement dated September 30, 1992 ("License Agreement," attached as Exhibit A and incorporated by reference) and under certain United States patents related to the License Agreement.

II. PARTIES

2. Plaintiff LIGAND PHARMACEUTICALS INCORPORATED (hereinafter "Ligand" or "Plaintiff") is a Delaware corporation with its principal place of business at 10275 Science Center Drive San Diego, California 92121.

3. Ligand was incorporated in 1987 and since then has been engaged in, *inter alia*, the research and development of drugs for various diseases and disorders. Ligand currently has less than sixty (60) employees.

4. Defendant THE ROCKEFELLER UNIVERSITY (hereinafter "Rockefeller" or "Defendant") is a New York not-for-profit corporation with its principal place of business at 1230 York Avenue, New York, New York 10021.

5. Rockefeller is a university periodically engaged in research and development. Rockefeller currently has 69 heads of laboratories, 200 research and clinical scientists, 350 postdoctoral investigators, 1,050 support staff, 150 Ph.D. students, 50 M.D.-Ph.D. students and 960 alumni according to the Rockefeller website.

6. NEW YORK UNIVERSITY ("NYU") is a New York not-for-profit corporation with its principal place of business at 70 Washington Square S, New York, New York 10012.

7. NYU is a university periodically engaged in research and development. NYU is not a party to the License Agreement or this lawsuit, but in the past it has received payments due to it under the License Agreement.

III. JURISDICTION AND VENUE

8. This Court has personal jurisdiction over Defendant Rockefeller by virtue of its presence and activities in the state of California, including but not limited to entering into

the License Agreement, as rights granted by the License Agreement were to be used in this judicial district, its past ownership interest in Ligand (located in this judicial district) under the License Agreement, as well as activities of Dr. James E. Darnell ("Darnell") in performing services in this judicial district under a Professional Services Agreement ("Services Agreement") dated September 30, 1992.

9. NYU is not being joined in this lawsuit for the following reasons. It is not a party to the License Agreement. Its interests under the License Agreement are subordinate to those of Rockefeller and, on information and belief, those interests are adequately protected by Rockefeller. Finally, Rockefeller, not NYU, is the owner of any intellectual property rights licensed under the License Agreement.

10. This Court has subject matter jurisdiction pursuant to 28 U.S.C. §§ 1332, 1338 and 2201.

11. Venue is proper in this judicial district pursuant to 28 U.S.C. § 1391(a) and (c).

IV. TECHNOLOGY

12. Since its inception, and prior to entering into the License Agreement with Rockefeller, Ligand has been actively involved in small molecule drug discovery. For example, Ligand owns intracellular receptor ("IR") technology that relates to families of transcription factors that change cell function by selectively turning on or off specific genes in response to circulating signals that act on cells. Ligand developed (and/or in-licensed from one or more sources other than Rockefeller) certain IR-based transcriptional assays to screen candidate drugs.

13. Thrombopoietin ("TPO") is a peptidyl hormone that activates a signaling cascade in a cell by binding to a receptor on a cell surface. Once bound by TPO, the cell surface receptor initiates a signaling cascade from the cell surface to the nucleus, where specific genes are selectively turned on in response to TPO. This gene regulation is mediated by transcription factors activated by the TPO signaling cascade and has a major effect on cell fate decisions by regulating cell proliferation and differentiation.

14. Ligand developed cell-based assays to screen candidate TPO mimics. These assays included cell proliferation and cell differentiation assays, as well as transcriptional assays. The transcriptional assays developed by Ligand to screen candidate TPO mimics were analogous to the transcriptional assays developed for Ligand's IR program.

15. The transcriptional assays involved use of a reporter construct with produces a signal in response to activated transcription factors in the cell.

16. Ligand's assays were used to discover and develop new drugs that mimic the action of TPO and may be useful in the treatment of a wide variety of diseases and disorders.

V. FACTUAL BACKGROUND

17. Darnell served on Ligand's Scientific Advisory Board for several years and visited with Ligand scientists at Ligand's facilities and elsewhere in San Diego many times in connection with the License Agreement and/or the Services Agreement.

18. On information and belief, at all times relevant here to, Darnell acted in conjunction with Rockefeller and had authority to act on behalf of Rockefeller to fulfill Rockefeller's obligations under the License Agreement.

19. After negotiations between the parties, Ligand executed two separate agreements on September 30, 1992, the License Agreement with Rockefeller and the Services Agreement with Darnell.

20. The License Agreement was generally directed to the licensing of "Licensed Patent Rights" and "Technical Information" relating to peptidyl hormone mediated gene expression.

21. The Licensed Patent Rights are defined in Section 1.3 of the License Agreement to be patent applications identified in Exhibit A to the License Agreement, related "divisionals, continuations, continuations-in-part, reissues, renewals, foreign counterparts, extension or additions," and any patents which may issue thereon. (Section 1.3, License Agreement).

22. Rockefeller is the identified assignee of United States patents, including: U.S. Pat. No. 6,605,442; U.S. Pat. No. 5,976,835; U.S. Pat. No. 6,013,475; U.S. Pat. No.

6,030,808; U.S. Pat. No. 6,338,949; U.S. Pat. No. 6,124,118; U.S. Pat. No. 7,060,682; U.S. Pat. No. 5,716,622; U.S. Pat. No. 5,883,228; U.S. Pat. No. 6,030,780; U.S. Pat. No. 6,720,154; U.S. Pat. No. 7,115,567; U.S. Pat. No. 6,960,647; and U.S. Pat. No. 7,211,655 ("Rockefeller Patents" attached as Exhibits B through O), which all either claim priority back to the patent applications listed in Exhibit A to the License Agreement or relate to what Rockefeller argues is Technical Information under the License Agreement.

23. Technical Information is defined in Section 1.4 of the License Agreement to include "technical data, information processes, materials and know-how, whether or not patentable" relating to peptidyl mediated gene expression that is owned by Rockefeller and was developed as of the effective date of the License Agreement or during the next five (5) years. (Section 1.4, License Agreement).

24. The License Agreement between Ligand and Rockefeller contemplated that certain of the intellectual property of Rockefeller might be used by Ligand in development of new pharmaceutical agents. (Sections 2.4 and 2.5, License Agreement). Nothing in the License Agreement prohibited Ligand from developing processes and products relating to cell-based assays to screen candidate drugs independent of Rockefeller's intellectual property, as Ligand had done previously with its IR technology.

25. Independent of the rights acquired under the License Agreement, on December 29, 1994, Ligand entered into a Research Development and License Agreement ("GSK License") with SmithKline Beecham Corporation, now GlaxoSmithKline ("GSK"). The GSK License relates to a joint research and development effort by Ligand and GSK directed to discovery of small molecule compounds which act as modulators of certain HEMATOPOIETIC GROWTH FACTORS (including TPO, as defined in Section 1.17 of the GSK License) and to develop pharmaceutical products from such compounds.

26. On information and belief, Rockefeller has been aware of the GSK License since it was signed by Ligand and GSK in 1994.

27. Under the RESEARCH PROGRAM as defined in the GSK License, a cell-based high throughput screen was developed by Ligand to help identify at least one

potentially useful drug known as eltrombopag or PROMACTA® and a back-up thereto known as SB-559448 ("GSK Products"). Under the GSK License, GSK has paid Ligand milestone payments amounting to $8 million for achieving certain milestones under the GSK License.

28. GSK has made significant progress toward gaining approval for at least one of the GSK Products through the regulatory process before the Food and Drug Administration.

29. As early as October 2003, Rockefeller became specifically aware of the GSK Products and inquired about and demanded payment from Ligand under the License Agreement for what Rockefeller alleged were uses of its Licensed Patent Rights or Technical Information covered by the License Agreement.

30. Ligand disputes that the GSK Products are subject to payments under the License Agreement.

31. Section 2.5 of the License Agreement obligates Ligand to pay Rockefeller only under certain circumstances. The payments described in Section 2.5 generally are twenty five per cent (25%) of payments received from third parties by Ligand if those payments were to secure the right to use Technical Information or the right to sell Products or Processes.

32. The GSK Products are not Products as the term "Product" is defined under Section 1.5 of the License Agreement. They do not embody or use any invention described or claimed in the Licensed Patent Rights. Furthermore, Technical Information was not essential to their discovery or development. GSK's payments to Ligand are not and will not be to secure any Rockefeller rights that would otherwise prevent GSK from selling the GSK Products. Rockefeller does not own any Licensed Patent Rights or Technical Information that GSK would need to sell the GSK Products. Thus, no payments are due to Rockefeller under the License Agreement.

33. Rockefeller has alleged the GSK Products embody or use one or more invention(s) described or claimed in the Licensed Patent Rights. In order to qualify as an invention in a claim of an issued patent, however, the alleged invention must be defined by a claim that is valid and enforceable.

34. Section 11.2 of the License Agreement provides that Ligand shall have the right to terminate any license grant at any time upon ninety days written notice.

35. On August 9, 2007, pursuant to Section 11.2, Ligand sent by facsimile and U.S. Mail a notice to Rockefeller of its intent to terminate the License Agreement. Pursuant to Section 11.2, the termination was effective under the License Agreement ninety days thereafter or on November 7, 2007.

36. Since termination of the License Agreement under Section 11.2, Rockefeller has claimed that the License Agreement was not terminated. Rockefeller contends that 25% of past and future payments related to GSK Products received by Ligand must be shared with Rockefeller.

37. The parties entered into a tolling agreement that contemplated the parties would try to resolve the controversy without the need for litigation. The tolling agreement expired on March 3, 2008. Rockefeller's communications prior to March 3, 2008, including their refusal to extend the tolling agreement and their specific threat of filing a lawsuit against Ligand at the expiration of the tolling agreement, have made Ligand reasonably afraid that it will be sued by Rockefeller on these issues today or within the next few days.

VI. FIRST CLAIM FOR RELIEF – DECLARATORY JUDGMENT SCOPE OF LICENSED PATENT RIGHTS

38. Ligand incorporates by reference as though fully set forth herein paragraphs 1 through 37 of this Complaint.

39. The License Agreement between Ligand and Rockefeller provides for, among other things, a license of Licensed Patent Rights. (Section 2.1, License Agreement).

40. Rockefeller has alleged that the Rockefeller Patents are included within the Licensed Patent Rights and also that the GSK Products or their use embody or employ the Licensed Patent Rights.

41. Applying the plain meaning of the words of the License Agreement, the GSK Products and their use do not embody or employ any invention described or claimed in the Licensed Patent Rights.

42. An actual controversy exists between Rockefeller and Ligand as to whether or not the GSK Products or their use embody or employ Licensed Patent Rights, whether or not the GSK Products or their use embody or employ any invention described or claimed in the Rockefeller Patents and whether or not the payments Rockefeller is demanding under the License Agreement are in fact due.

43. Even if the GSK Products embody or use an invention merely described in the Rockefeller Patents, the patent laws of the United States protect only inventions defined by valid and enforceable claims and there is an actual controversy as to whether or not any claim of the Rockefeller Patents is valid for failure to comply with any one of 35 USC §§ 101 et seq.

44. On information and belief, Rockefeller has filed one or more patent applications for the purpose of claiming the GSK Products are subject to payments under the License Agreement, and Rockefeller did so with knowledge that no valid patent should issue. There is an actual controversy as to whether the GSK Products or their use embody or employ any invention described or claimed in any pending patent application and whether any such patent application filed after learning of the GSK Products was filed in good faith under the License Agreement.

VII. SECOND CLAIM FOR RELIEF – DECLARATORY JUDGMENT SCOPE OF TECHNICAL INFORMATION

45. Ligand incorporates by reference as though fully set forth herein paragraphs 1 through 44 of this Complaint.

46. The License Agreement between Ligand and Rockefeller provides for, among other things, a license of Technical Information of Rockefeller. (Section 2.1, License Agreement).

47. Rockefeller alleges that Technical Information was essential to the discovery or development of the GSK Products.

48. Ligand, relying on the plain meaning of the License Agreement, alleges that Technical Information was not used in the discovery or development of the GSK Products.

Ligand further alleges under Section 1.4 of the License Agreement Technical Information must be owned by Rockefeller and existing or capable of description in a tangible form and must have been developed in the laboratory of Darnell or of David Levy of NYU as of September 30, 1992 or by Darnell at his laboratory on or before five years from September 30, 1992 or by September 30, 1997. The GSK Products were not developed using Technical Information but rather used either publicly known information, information known or discovered by Ligand and/or GSK, or information received from third parties.

49. An actual controversy exists between Rockefeller and Ligand as to whether or not Technical Information was essential to the discovery or development of the GSK Products.

VII. THIRD CLAIM FOR RELIEF – DECLARATORY JUDGMENT TERMINATION

50. Ligand here incorporates by reference as though fully set forth herein paragraphs 1 through 49 of this Complaint.

51. Rockefeller relies on Section 11.3 of the License Agreement in asserting that, absent a material breach, the "Agreement" cannot be terminated.

52. Ligand claims, in the alternative, that the notice dated August 9, 2007 either terminated the License Agreement in its entirety, subject only to certain specified rights which survived termination, or to the extent any different, terminated all then existing license rights, again subject only to any rights that might survive termination.

53. An actual controversy exists between Rockefeller and Ligand as to whether or not the License Agreement has been terminated and as to the nature of the rights terminated.

VIII. DEMAND FOR JUDGMENT

WHEREFORE, Plaintiff requests that:

1. This Court enter a judgment declaring the GSK Products do not embody any invention(s) described or claimed in the Licensed Patent Rights and that the use of the GSK Products do not employ any invention(s) described or claimed in the Licensed Patent Rights;

2. This Court enter a judgment declaring that Technical Information was not essential to the discovery or development of the GSK Products;

3. This Court enter a judgment declaring that Ligand is not liable for any additional payments under the License Agreement beyond those that have already been made;

4. This Court enter a judgment declaring that the License Agreement was terminated as of November 7, 2007 and that subsequent to termination of the License Agreement, Ligand is not liable for any future payments under the License Agreement;

5. Plaintiff be awarded costs, attorneys' fees and other relief, both legal and equitable, to which it may be justly entitled;

6. Plaintiff be awarded relief under 28 U.S.C. § 2202; and

7. Plaintiff be awarded such other and further relief as this Court deems proper.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: 3/3/08

By: Darrell Olson
Darrell Olson (signature via facsimile)

Attorneys for Plaintiff
LIGAND PHARMACEUTICALS INCORPORATED

EXHIBIT A

LICENSE AGREEMENT

AGREEMENT made as of the 30th day of September, 1992 ("Effective Date") by and between LIGAND PHARMACEUTICALS INCORPORATED ("Ligand"), a corporation organized and existing under the laws of the State of Delaware, having a principal place of business at 9393 Towne Centre Drive, San Diego, California 92121, and THE ROCKEFELLER UNIVERSITY ("Rockefeller"), a nonprofit education corporation organized and existing under the laws of the State of New York, having an office at 1230 York Avenue, New York, New York 10021.

W I T N E S S E T H:

WHEREAS, Dr. James Darnell and his colleagues at Rockefeller and at NYU have developed valuable technology and know-how relating to peptidyl hormone mediated gene expression, including application for patents thereon, which constitutes core technology to be licensed hereunder;

WHEREAS, NYU has assigned to Rockefeller its rights to the core technology;

WHEREAS, Rockefeller has the right to grant exclusive license rights with respect to such core technology and to future developments relating thereto made at Rockefeller in the manner described herein; and

WHEREAS, Ligand wishes to obtain the exclusive license rights described herein for commercial development and application;

NOW, THEREFORE, in consideration of the mutual benefits to be derived hereunder, the parties hereto agree as follows:

1. Definitions

The following terms will have the meanings assigned to them below when used in this Agreement.

1.1 "Party" shall mean either Ligand or Rockefeller and "Parties" shall mean Ligand and Rockefeller.

1.2 "Affiliate" shall mean a corporation or other entity which directly or indirectly controls, is controlled by or under common control with Ligand. An entity shall be regarded as in control of another if it owns, or directly or indirectly controls, at least 50% of the voting stock or other ownership interest of the other entity, or if it directly or indirectly possesses the power to direct or cause the direction of the management and policies of the other entity by any means whatsoever.

1.3 "Licensed Patent Rights" shall mean

(a) the patent application(s) set forth on Exhibit "A" attached hereto and all patents which may issue thereon;

(b) the patent applications which are divisionals, continuations, continuations-in-part, reissues, renewals, foreign counterparts, extensions or additions of the patents and/or applications described in (a) and (b) of this paragraph 1.3, and all patents which may issue thereon;

(c) and all other patent applications, and patents issuing thereon, filed to cover Technical Information, including divisionals, continuations-in-part, reissues, renewals, foreign counterparts, extensions or additions and patents which may issue thereon.

1.4 "Technical Information" shall mean any and all technical data, information processes, materials and know-how,

whether or not patentable, owned by Rockefeller and existing or capable of description in a tangible form relating to peptidyl hormone mediated gene expression (a) developed in the laboratory of Dr. James Darnell of Rockefeller or Dr. David Levy of NYU as of the Effective Date and (b) which is subsequently developed at Rockefeller in the laboratory of Dr. James Darnell during the period ending five (5) years from the Effective Date.

1.5 "Product" shall mean any product which embodies or the use of which employs any invention(s) described or claimed in Licensed Patent Rights or for which Technical Information was essential to the discovery or development thereof.

1.6 "Process" shall mean any process which embodies or the practice of which employees any invention(s) described or claimed in Licensed Patent Rights or for which Technical Information was essential to the discovery or development thereof.

1.7 "Territory" shall mean the entire world.

1.8 "Net Sales" shall mean, in the case of sales to non-Affiliates, the invoiced price by Ligand or Affiliates less (a) customary trade quantity and cash discounts actually allowed and taken; (b) allowances actually given for returned, rejected or recalled Products; actual charges for bad debts; (c) freight and insurance if included in the price; government mandated rebates; and (d) value added tax, sales, use or turnover taxes, excise taxes, and custom duties included in the invoiced price.

2. License Rights

2.1 Rockefeller hereby grants to Ligand a sole exclusive license, including the right to grant royalty bearing sublicenses under terms consistent with this Agreement under Licensed Patent Rights and Technical Information, to make, have made, use and sell Products or practice Processes in any country of

the Territory, except to the extent that Rockefeller's right to do so may be limited under the provisions of the following:

(a) 35 United States, Section 201 et seq., and regulations and rules promulgated thereunder, or

(b) other applicable laws or regulations of the United States;

Provided only that Rockefeller is satisfied that the licensee is making a substantial and good faith effort to achieve practical application of the subject invention and its public use, Rockefeller agrees to use reasonable and proper efforts to extend exclusivity of the license consistent with the aforesaid U.S. government rights and policies should U.S. government action limit such exclusivity.

2.2 In consideration of the Ligand stock to be issued to Rockefeller and NYU as described in Section 2.3 and the cash payments to be made pursuant to Section 2.3, the license to Ligand under Section 2.1 shall be deemed to be fully paid up for research purposes including for the purposes of research done by Ligand or a Ligand sublicensee or collaboratively with a third party to the extent that the third party payments to Ligand do not exceed its fully burdened costs for performance of such research and development.

2.3 On the Effective Date, Ligand shall transfer to Rockefeller and NYU collectively a total of 150,000 shares of Series G Preferred Stock pursuant to Stock Transfer Agreements of even date herewith, 100,000 shares of which will vest on the Effective Date and 50,000 shares of which will vest in two installments of 25,000 shares on the first and second anniversaries hereof unless this Agreement is sooner terminated as provided herein. On the Effective Date, Ligand will also grant Rockefeller and NYU collectively, five year, net issuance warrants to purchase

a total of 100,000 shares of Ligand common stock vesting and exercisable as follows:

(i) a total of 50,000 shares vesting at the third anniversary of the Effective Date and exercisable at $14.00 per share; and

(ii) a total of 50,000 shares vesting at the fourth anniversary of the Effective Date exercisable at the fair market value on the vesting date.

As further consideration, Ligand will make cash payments to Rockefeller and NYU pursuant to the following schedule:

(a) On the Effective Date;

Rockefeller	$45,000
NYU	$ 5,000

(b) $67,500 to Rockefeller and $7,500 to NYU when the current Technical Information is successfully transferred to Ligand as described in Section 5;

(c) $67,500 to Rockefeller and $7,500 to NYU on each of the 1st - 4th anniversaries of the Effective Date.

2.4 Ligand will pay a royalty of five percent (5%) of its Net Sales of Products and on its net revenues, i.e., gross revenues less fully burdened costs, received from performance of Processes for a third party. The royalty shall be paid for a term which is the longer of ten (10) years or, on a country by country basis, expiration of the last patent in the Licensed Patent Rights having a claim which reads on the Product or Process or a method of making or using the Product or Process. Only one royalty will be owed on a Product or Process in the circumstance where the Product or Process is covered by multiple claims in the Licensed Patent Rights. Royalty payments made under this Section 2.4 and under

Section 2.5 shall be made to Rockefeller and NYU in the ratio ninety percent (90%) to Rockefeller and ten percent (10%) to NYU.

2.5 In the case of payments made to Ligand by a third party to secure the right to use Technical Information or to sell Products or Processes, Ligand will pay to Rockefeller and NYU twenty-five percent (25%) of the payments made to Ligand by the third party; provided, however, that in the situation where the payment to Ligand is based on the third party's revenues arising from sale of a Product or use of a Process, then Ligand shall pay to Rockefeller and NYU the lesser of twenty-five percent (25%) of the payment received from the third party or a royalty calculated pursuant to Section 2.4 by treating the third party's sales of such Products and Processes as Ligand sales. Payments by a third party to Ligand to purchase equity in Ligand and to fund research at Ligand which do not generate net revenue as defined in Section 2.4 shall not be subject to sharing under this Section 2.5.

2.6 In the event Ligand is required to make payments to a third party to use Technical Information, it shall be entitled to credit fifty percent (50%) of that payment against any royalty owed under this Agreement but in no event may it reduce a payment owed by more than fifty percent (50%).

2.7 Ligand will diligently seek to develop Products and/or Processes using or based on Technical Information. Ligand shall be deemed to have met its diligence obligations during the first five (5) years of the Agreement if, in the aggregate, Ligand, its Affiliates, licensees and research collaborators expend at least $4,000,000 directed toward the development of Products and Processes and support at least ten (10) full time scientist equivalents in support of that effort.

3. Patents

3.1 The Company agrees to reimburse Rockefeller for all amounts expended prior to the date hereof for the preparation, filing, prosecution and maintenance of Licensed Patent Rights licensed to the Company pursuant to Section 2.1 of this Agreement, said amount being $20,791.18 as of September 8, 1992.

3.2 The Company shall continue to reimburse Rockefeller for such reasonable additional filing, prosecution, and maintenance costs as shall be incurred on each such patent application or patent licensed hereunder during the term of such license.

3.3 Rockefeller shall select qualified independent patent counsel reasonably satisfactory to Ligand to file and prosecute all patent applications included in Licensed Patent Rights, including divisionals, continuations, continuations-in-part, reissues, and foreign counterparts. Such counsel shall regularly meet and/or consult with Ligand and/or its designated officers and counsel to keep them advised of the status of patent matters in the normal course. Patent counsel shall be instructed not to file any papers without giving Ligand ample time and opportunity to review and comment. Ligand shall be entitled to determine the countries in which it wishes to obtain and maintain patent protection under this Agreement and shall be free, at any time and at its sole option, to abandon patent prosecution or maintenance in any country of the Territory.

3.4 Ligand shall promptly advise Rockefeller of any decision not to finance the preparation, filing, prosecution or maintenance of any patent application or patent licensed hereunder in adequate time to allow Rockefeller, at its own cost, to effectuate such preparation, filing, prosecution, or maintenance if it desires to do so; and Ligand shall, at the request of Rockefeller, take whatever steps may be necessary to return to Rockefeller all rights which Ligand may have with respect to the

applicable Licensed Patent Rights and Technical Information which it proposes to abandon.

Nothing herein is intended or shall be construed as obligating Rockefeller to apply for any U.S. or foreign patents at its own expense, or to defend, enforce, or support any patent or patent application which may be included in Licensed Patent Rights to which it has granted license rights to Ligand; provided, however, that Rockefeller will cooperate with Ligand in Ligand's activity in applying for U.S. or foreign patents or in the defense or enforcement of Licensed Patent Rights.

Nothing herein is intended or shall be construed as obligating Ligand to maintain its license with respect to any patent or application licensed hereunder and to finance the preparation, filing, prosecution or maintenance of any patent application in any bounty or jurisdiction in which it believes it is not in the best business interests.

3.5 Ligand shall have the right to institute patent infringement proceedings against third parties based on any Licensed Patent Rights licensed hereunder. If Ligand does not institute infringement proceedings against such third parties, Rockefeller shall have the right but not the obligation, to institute such proceedings. Within thirty (30) days after notice of its intention to commence such proceedings given to Ligand and provided that Ligand does not, within such thirty (3) day period, institutes its own proceedings, Rockefeller may institute such proceedings. The expenses of such proceedings, including lawyers' fees, shall be borne by the Party instituting suit. Each Party shall execute all necessary and proper documents and take all other appropriate action to allow the other Party to institute and prosecution such proceedings. Any award paid by third parties as a result of such proceedings (whether by way of settlement or otherwise) shall first be applied toward reimbursement for the legal fees and expenses incurred, and the excess, if any, shall be

shared on a pro rata basis based on the expenses incurred by each party.

3.6 Should Ligand decide at any time during the term hereof that it will no longer commercially pursue the development of any invention licensed hereunder, Ligand shall promptly notify Rockefeller of its decision and, upon request from Rockefeller, shall take whatever steps are necessary to assure reversion to Rockefeller of all rights to that invention.

3.7 Ligand shall assume the responsibility at its own expense, and using counsel of its choosing, to defend against claims of patent infringement arising from the making, using, or selling of Products and Processes.

4. Payments and Reports

4.1 Within forty-five (45) days of the end of each calendar quarter during the term of this Agreement, beginning with the first quarter in which the obligation to make a payment to Rockefeller arises, Ligand shall submit to Rockefeller and NYU a report in writing setting forth the net revenues (revenues less Fully Burdened Costs) earned from the performance of a Process and the Net Sales of Products, and payments to Ligand which are subject to sharing with Rockefeller and NYU. The report shall include a calculation of the payments owed to Rockefeller and NYU arising therefrom and shall be accompanied by payment to Rockefeller and NYU in the full amount thereof.

4.2 Ligand shall keep adequate records in sufficient detail to enable the payments due from Ligand hereunder to Rockefeller and NYU to be determined, and permit said records to be inspected at any time during regular business hours at its principal place of business by an independent certified public accountant appointed by Rockefeller, or Rockefeller and NYU together but not NYU alone, for this purpose and who is reasonably

acceptable to Ligand. The accountant shall be required to enter into a confidentiality agreement with Ligand substantially in the form of the provisions contained in Article 5 herein and shall only report to Rockefeller, and NYU if a joint audit is done, the discrepancy, if any, between the amount owed by Ligand for the audited period and the amount actually paid and discrepancies in the method of calculating Fully Burdened Costs. Ligand shall maintain such records for a minimum of three years. No more than one such audit shall be requested per calendar year. Rockefeller, or Rockefeller and NYU if a joint audit, shall bear the cost of any such audit; provided, however, that where the auditor determines that the payments owed for an audit period exceeds that paid to Rockefeller and NYU by Ligand by more than ten (10) percent, the reasonable cost of the audit shall be borne by Ligand.

5. Technical Information Transfer

Rockefeller will diligently cooperate with Ligand to transfer Technical Information to Ligand. Transfer of current Technical Information will be deemed to have successfully occurred for the purposes of Section 2.3 when Rockefeller has transferred to Ligand, and Ligand has successfully expressed, functional proteins from the clones of the genes specifically described in the applications for United States Patents identified in Exhibit "A".

6. Confidentiality

6.1 The Parties contemplate that during the course of their relationship arising under this Agreement it may be necessary to provide the other with confidential information to facilitate the performance of their obligations pursuant to this Agreement. The Parties agree, therefore, that information received from the other which is in writing and identified as confidential, or if disclosed orally, is confirmed in writing and designated confidential, shall be maintained in confidence and that reasonable

and prudent practices shall be followed to maintain the information in confidence, including, where necessary, obtaining written confidentiality agreements from employees not already bound by such agreements who have access to the confidential information. Information received in confidence shall be used by a party only for the purpose of and in connection with its performance of this Agreement. The obligation to maintain information in confidence shall survive this Agreement or termination thereof for any reason for a period of five (5) years thereafter. However, a party shall not be obliged to maintain information in confidence which it can show by written documentation: (a) to have been publicly known prior to submission to it; (b) to have been known or available to it prior to submission by the other party; (c) to have become publicly known without fault on its part subsequent to submission by the other party; (d) to have been received by it from a third party legally having possession of the information without obligations of confidentiality; (e) to be required to be disclosed pursuant to order of any court or governmental agency having jurisdiction thereof after notice to the other party sufficient to afford it an opportunity to intervene in the proceeding where disclosure is required; and (f) to be necessarily revealed in the course of marketing any Product or Process which is licensed hereunder.

7. Academic Freedom

Rockefeller and Ligand recognize the traditional freedom of all scientists to publish and present promptly the results of their research. Rockefeller and Ligand also recognize that exclusive patent rights can be jeopardized by public disclosure prior to the filing of suitable patent applications. Therefore, Rockefeller will assure that each proposed publication concerning any technology described in Licensed Patent Rights or which may constitute an Improvement thereof, before submission to a publisher, will be submitted to Ligand for review in connection with preservation of exclusive patent rights. Ligand shall have

thirty (30) days in which to review the publication, which may be extended for an additional thirty (30) days when Ligand provides substantial and reasonable need for such extension. By mutual agreement, this period may be further extended for not more than an additional three (3) months. Ligand will allow for simultaneous submission of the publication to the publisher and Ligand, where appropriate. Any publication by Ligand personnel will also be subject to similar pre-review before publication. Scientists at Rockefeller and Ligand will be expected to treat matters of authorship in a proper collaborative spirit, giving credit where it is due and proceeding in a manner which fosters cooperation and communication.

8. Warranty

8.1 Rockefeller warrants that it has the right to grant to the full extent thereof the license granted Ligand hereunder and that it has and will discharge their duty of disclosure to the United States Patent and Trademark Office.

8.2 EXCEPT AS WARRANTED IN THE PRIOR SECTION 8.01, ROCKEFELLER MAKES NO WARRANTY, EXPRESS OR IMPLIED, INCLUDING ANY EXPRESS OR IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR PARTICULAR PURPOSE.

9. Publicity

Ligand will not use directly or by implication the name of Rockefeller, or the name of any member of the faculty or staff of Rockefeller, or any unpublished information or data relating to the investigation for any business, promotional, commercial or other purpose, without the prior written approval of Rockefeller and the faculty or staff member involved; except Ligand may use and disclose such names in its internal communications or in any required governmental reports and filings upon prior disclosure and consultation with Rockefeller, as appropriate.

10. Product Liability

Ligand agrees to indemnify and hold harmless Rockefeller, its trustees, officers, agents, faculty, employees, and students from any and all liability arising from injury or damage to persons or property resulting directly or indirectly from Ligand's acquisition, use, manufacture, or sale of any Product covered by Licensed Patent Rights or Technical Information licensed hereunder. Ligand further agrees, so long as it is selling any Product, to obtain and maintain in force product liability insurance coverage in amounts reasonably satisfactory to Rockefeller, as appropriate to the risk as determined by reference to reliable standards in the industry, such insurance to specifically name Rockefeller as an additional insured.

11. Termination

11.1 The licenses herein granted shall continue for the full term of any patents licensed hereunder as the same or the effectiveness thereof may be extended by any governmental authority, rule or regulation applicable thereto.

11.2 Ligand shall have the right to terminate any license grant at any time upon ninety (90) days' prior written notice to Rockefeller, provided, however, that termination shall not affect Rockefeller's and NYU's rights and privileges as a stockholder of Ligand or their ownership of any vested shares of Ligand.

11.3 Any Party may terminate this Agreement in the event of a material breach by the other party, provided only that the offending Party is given notice of the breach and a reasonable time, not to exceed sixty (60) days, in which to cure such breach.

11.4 Any termination of this Agreement and of any option and/or license granted hereunder shall also terminate any applicable sublicense thereunder.

11.5 The Parties acknowledge that Ligand's right to the future developments made at Rockefeller in the laboratory of Dr. James Darnell are an important element of this Agreement. Therefore, in the event that Dr. Darnell for health reasons or otherwise ceases to actively conduct research at Rockefeller as a full time member of the faculty, then Ligand can, without loss of rights under the Agreement, terminate the making of anniversary cash payments under Section 2.3.

12. Notices

Any notice required to be given pursuant to this Agreement shall be made by personal delivery or, if by mail, then by registered or certified mail, return receipt requested, with postage and fees prepaid, by one Party to the other Party at the addresses noted below.

In the case of Ligand, notice should be sent to:

Ligand Pharmaceuticals Incorporated
9393 Towne Centre Drive, Suite 100
San Diego, CA 92121
Attn: General Counsel

In the case of Rockefeller, notice should be sent to:

The Rockefeller University
1230 York Avenue
New York, NY 10021
Attn: Office of the General Counsel

13. Law to Govern

This Agreement shall be interpreted and governed in accordance with the laws of the State of New York.

14. No Partnership

This Agreement shall not constitute a partnership or a joint venture, and neither Party may be bound by the other to any

contract, arrangement or understanding except as specifically stated herein.

15. No Waiver

The failure of either party to enforce at any time any of the provisions of this Agreement, or any rights in respect thereto, or to exercise any election herein provided, shall in no way be considered to be a waiver of such provisions, rights or elections, or in any way to affect the validity of this Agreement. Exercise by either party any of its rights herein or any of its elections under the terms or covenants herein shall not preclude either party from exercising the same or any other rights in this Agreement, irrespective of any previous action or proceeding taken by either party hereunder.

16. Severability

If any provision of this Agreement is judicially determined to be void or unenforceable, such provision shall be deemed to be severable from the other provisions of this Agreement which shall remain in full force and effect. Either Party may request that a provision otherwise void or unenforceable be reformed so as to be valid and enforceable to the maximum extent permitted by law.

17. Assignment

This Agreement may not be assigned by either Party without the prior written consent of the other, which consent shall

not be unreasonably withheld except that Ligand may assign this Agreement to a successor entity in the case of a merger, acquisition or other reorganization.

18. Resolution of Dispute

The Parties agree that in the event of a dispute between them arising from concerning, or in any way relating to this Agreement, the Parties shall undertake good faith efforts to resolve the same amicably between themselves.

19. Force Majeure

The Parties shall not be liable in any manner for failure or delay in fulfillment of all or party of this Agreement, directly or indirectly caused by acts of God, governmental orders or restrictions, war, war-like condition, revolution, riot, looting, strike, lockout, fire, flood or other similar or dissimilar causes or circumstances beyond the non-performing Party's control. The non-performing Party shall promptly notify the other Party of the cause or circumstance and shall recommence its performance of its obligations as soon as practicable after the cause or circumstance ceases.

20. Entire Understanding

This Agreement, together with the Exhibits hereto, and the further documents and agreements executed in connection with the transactions contemplated hereby constitute the entire agreement between the Parties and supersedes all prior

understandings and agreements by the Parties with respect to the subject matter hereof.

IN WITNESS WHEREOF, the Parties have caused this Agreement to be duly executed as of the day and year first above written.

THE ROCKEFELLER UNIVERSITY

By _____

Title President

LIGAND PHARMACEUTICALS INCORPORATED

By _____

Title President and CEO

EXHIBIT "A"

U. S. PATENT APPLICATIONS

1. TITLE: "RECEPTOR RECOGNITION FACTOR AND METHODS OF USE THEREOF"
INVENTORS: Darnell and Levy
SERIAL NO.: 07/613,326
FILED: November 14, 1990

2. TITLE: "RECEPTOR RECOGNITION FACTORS, PROTEIN SEQUENCES AND METHODS OF USE THEREOF"
INVENTORS: Darnell, Schindler and Fu
SERIAL NO.: 07/854,296
FILED: March 19, 1992

US006605442B1

(12) **United States Patent**
Darnell, Jr. et al.

(10) **Patent No.: US 6,605,442 B1**
(45) **Date of Patent: Aug. 12, 2003**

(54) **METHODS OF TESTING DRUGS OR AGENTS THAT MODULATE THE ACTIVITY OF RECEPTOR RECOGNITION FACTORS**

(75) Inventors: **James E. Darnell, Jr.**, Larchmont, NY (US); **Christian W. Schindler**, New York, NY (US); **Xin-Yuan Fu**, Forrest Hills, NY (US); **Zilong Wen**, New York, NY (US); **Zhong Zhong**, New York, NY (US)

(73) Assignee: **The Rockefeller University**, New York, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **08/212,185**

(22) Filed: **Mar. 11, 1994**

Related U.S. Application Data

(63) Continuation-in-part of application No. 08/126,588, filed on Sep. 24, 1994, now abandoned, and a continuation-in-part of application No. 08/126,595, filed on Sep. 24, 1994, now abandoned, each is a continuation-in-part of application No. 07/980,498, filed on Nov. 23, 1992, now abandoned, which is a continuation-in-part of application No. 07/854,296, filed on Mar. 19, 1992, now abandoned.

(51) **Int. Cl.**[7] **G01N 33/53**
(52) **U.S. Cl.** **435/7.21**; 435/7.1; 435/7.2
(58) **Field of Search** 530/350, 351; 436/501; 435/69.1, 70.1, 183, 172.1, 240.2, 7.21, 7.1, 7.2; 935/36

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,463,023 A * 10/1995 Schreiber et al. 530/327
5,618,678 A * 4/1997 Kelly et al. 435/7.21

OTHER PUBLICATIONS

X-F Fu et al., PNAS 87:8555:8559, Nov. 1990.*
Improta et al., 1994, Proc. Natl. Acad. Sci. USA 91:4776–80.
Shuai et al., 1994, Cell 76:821–28.
Eck et al., 1993, Nature 362:87–91.
Felder et al., 1993, Mol. Cell. Biol. 13:1449–55.
Khan et al., 1993, Proc. Natl. Acad. Sci. USA 90:6806–10.
Müller et al., 1993, EMBO J. 12:4221–28.
Müller et al., 1993, Nature 366:129–35.
Pearse et al., 1993, Proc. Natl. Acad. Sci. 90:4314–18.
Sadowski et al., 1993, Science 261:1739–44.
Shuai et al., 1994, Nature 366:580–83.
Shuai et al., 1993, Science 261:1744–46.
Songyang et al., 1993, Cell 72:767–78.
Watling et al., 1993, Nature 366:166–70.
Booker et al., 1992, Nature 358:684–87.
Fu, 1992, Cell 70:323–35.
Fu et al., 1992, Proc Natl. Acad Sci. USA 89:7840–43.
Overduin et al., 1992, Cell 70:697–704.
Schindler et al., 1992, Proc. Natl. Acad. Sci. USA 89:7836–39.
Schlinder et al., 1992, Science 257:809–13.
Shuai et al., 1992, Science 258:1808–12.
Velazqez et al., 1992, Cell 70:313–22.
Decker et al., 1991, Mol. Cell. Biol. 11:5147–53.
Decker et al., 1991, EMBO J. 10:927–32.
Kaplan et al., 1991, Nature 350:158–60.
Kessler and Levy, 1991, J. Biol. Chem. 266:23471–76.
Koch et al., 1991, Science 252:668–74.
Lew et al., 1991, Mol. Cell. Biol. 11:182–91.
McKendry et al, 1991, Proc. Natl. Acad. Sci. USA 88:11455–59.
Qureshi et al., 1991, J. Biol. Chem. 266:20594–97.
Fu et al., 1990, Proc. Natl. Acad. Sci. USA 87:8555–59.
Kessler et al., 1990, Genes Dev. 4:1753–65.
Lee et al., 1990, Mol. Cell. Biol. 10: 1982–88.
Levy and Darnell, 1990, The New Biologist 2:923–28.
Reich and Pfeffer, 1990, Proc. Natl. Acad. Sci. USA 87:8761–65.
Uzé et al., 1990, Cell 60:225–34.
Dale et al., 1989, Proc. Natl. Acad. Sci. USA 86:1203–7.
Decker et al., 1989, EMBO J. 8:2009–14.
Levy et al., 1989, Genes Dev. 3:1362–71.
Lew et al., 1989, Mol. Cell. Biol. 9:5404–11.
Aguet et al., 1988, Cell 55:273–80.
Levy et al., 1988, Genes Dev. 2:383–93.
Rudd et al., 1988, Proc. Natl. Acad. Sci. USA 85:5190–94.
Veillette et al., 1988, Cell 55:301–8.
Celis et al., 1987, Leukemia 1:800–13.
Leung et al., 1987, Nature 330:537–43.
Chodosh et al., 1986, Mol. Cell. Biol. 6:4723–33.
Friedman et al., 1984, Cell 38:745–55.

* cited by examiner

Primary Examiner—Lorraine Spector
(74) *Attorney, Agent, or Firm*—Klauber & Jackson

(57) **ABSTRACT**

Receptor recognition factors exist that recognizes the specific cell receptor to which a specific ligand has been bound, and that may thereby signal and/or initiate the binding of the transcription factor to the DNA site. The receptor recognition factor is in one instance, a part of a transcription factor, and also may interact with other transcription factors to cause them to activate and travel to the nucleus for DNA binding. The receptor recognition factor appears to be second-messenger-independent in its activity, as overt perturbations in second messenger concentrations are of no effect. The concept of the invention is illustrated by the results of studies conducted with interferon (IFN)-stimulated gene transcription, and particularly, the activation caused by both IFNα and IFNγ. Specific DNA and amino acid sequences for various human and murine receptor recognition factors are provided, as are polypeptide fragments of two of the ISGF-3 genes, and antibodies have also been prepared and tested. The polypeptides confirm direct involvement of tyrosine kinase in intracellular message transmission. Numerous diagnostic and therapeutic materials and utilities are also disclosed.

8 Claims, 45 Drawing Sheets

FIG. 1A

```
                                   1
                                  met ala gln trp glu met leu gln
ACTGCAACCCTAATCAGAGCCCAA          ATG GCG CAG TGG GAA ATG CTG CAG

     10                                                  20
asn leu asp ser pro phe gln asp gln leu his gln leu tyr ser
AAT CTT GAC AGC CCC TTT CAG GAT CAG CTG CAC CAG CTT TAC TCG

                              30
his ser leu leu pro val asp ile arg gln tyr leu ala val trp
CAC AGC CTC CTG CCT GTG GAC ATT CGA CAG TAC TTG GCT GTC TGG

     40                                                  50
ile glu asp gln asn trp gln glu ala ala leu gly ser asp asp
ATT GAA GAC CAG AAC TGG CAG GAA GCT GCA CTT GGG AGT GAT GAT

                              60
ser lys ala thr met leu phe phe his phe leu asp gln leu asn
TCC AAG GCT ACC ATG CTA TTC TTC CAC TTC TTG GAT CAG CTG AAC

     70                                                  80
tyr glu cys gly arg cys ser gln asp pro glu ser leu leu leu
TAT GAG TGT GGC CGT TGC AGC CAG GAC CCA GAG TCC TTG TTG CTG

                              90
gln his asn leu arg lys phe cys arg asp ile gln pro phe ser
CAG CAC AAT TTG CGG AAA TTC TGC CGG GAC ATT CAG CCC TTT TCC

    100                                                 110
gln asp pro thr gln leu ala glu met ile phe asn leu leu leu
CAG GAT CCT ACC CAG TTG GCT GAG ATG ATC TTT AAC CTC CTT CTG

                             120
glu glu lys arg ile leu ile gln ala gln arg ala gln leu glu
GAA GAA AAA AGA ATT TTG ATC CAG GCT CAG AGG GCC CAA TTG GAA

    130                                                 140
gln gly glu pro val leu glu thr pro val glu ser gln gln his
CAA GGA GAG CCA GTT CTC GAA ACA CCT GTG GAG AGC CAG CAA CAT

                             150
glu ile glu ser arg ile leu asp leu arg ala met met glu lys
GAG ATT GAA TCC CGG ATC CTG GAT TTA AGG GCT ATG ATG GAG AAG

    160                                                 170
leu val lys ser ile ser gln leu lys asp gln gln asp val phe
CTG GTA AAA TCC ATC AGC CAA CTG AAA GAC CAG CAG GAT GTC TTC

                             180
```

Session Name: rb

FIG. 1B

```
cys phe arg tyr lys ile gln ala lys gly lys thr pro ser leu
TGC TTC CGA TAT AAG ATC CAG GCC AAA GGG AAG ACA CCC TCT CTG

    190                                             200
asp pro his gln thr lys glu gln lys ile leu gln glu thr leu
GAC CCC CAT CAG ACC AAA GAG CAG AAG ATT CTG CAG GAA ACT CTC

                        210
asn glu leu asp lys arg arg lys glu val leu asp ala ser lys
AAT GAA CTG GAC AAA AGG AGA AAG GAG GTG CTG GAT GCC TCC AAA

    220                                             230
ala leu leu gly arg leu thr thr leu ile glu leu leu leu pro
GCA CTG CTA GGC CGA TTA ACT ACC CTA ATC GAG CTA CTG CTG CCA

                        240
lys leu glu glu trp lys ala gln gln gln lys ala cys ile arg
AAG TTG GAG GAG TGG AAG GCC CAG CAG CAA AAA GCC TGC ATC AGA

    250                                             260
ala pro ile asp his gly leu glu gln leu glu thr trp phe thr
GCT CCC ATT GAC CAC GGG TTG GAA CAG CTG GAG ACA TGG TTC ACA

                        270
ala gly ala lys leu leu phe his leu arg gln leu leu lys glu
GCT GGA GCA AAG CTG TTG TTT CAC CTG AGG CAG CTG CTG AAG GAG

    280                                             290
leu lys gly leu ser cys leu val ser tyr gln asp asp pro leu
CTG AAG GGA CTG AGT TGC CTG GTT AGC TAT CAG GAT GAC CCT CTG

                        300
thr lys gly val asp leu arg asn ala gln val thr glu leu leu
ACC AAA GGG GTG GAC CTA CGC AAC GCC CAG GTC ACA GAG TTG CTA

    310                                             320
gln arg leu leu his arg ala phe val val glu thr gln pro cys
CAG CGT CTG CTC CAC AGA GCC TTT GTG GTA GAA ACC CAG CCC TGC

                        330
met pro gln thr pro his arg pro leu ile leu lys thr gly ser
ATG CCC CAA ACT CCC CAT CGA CCC CTC ATC CTC AAG ACT GGC AGC

    340                                             350
lys phe thr val arg thr arg leu leu val arg leu gln glu gly
AAG TTC ACC GTC CGA ACA AGG CTG CTG GTG AGA CTC CAG GAA GGC

                        360
asn glu ser leu thr val glu val ser ile asp arg asn pro pro
AAT GAG TCA CTG ACT GTG GAA GTC TCC ATT GAC AGG AAT CCT CCT

    370                                             380
gln leu gln gly phe arg lys phe asn ile leu thr ser asn gln
CAA TTA CAA GGC TTC CGG AAG TTC AAC ATT CTG ACT TCA AAC CAG

                        390
lys thr leu thr pro glu lys gly gln ser gln gly leu ile trp
```

FIG. 1C

Session Name: rb

```
AAA ACT TTG ACC CCC GAG AAG GGG CAG AGT CAG GGT TTG ATT TGG

    400                                             410
asp phe gly tyr leu thr leu val glu gln arg ser gly gly ser
GAC TTT GGT TAC CTG ACT CTG GTG GAG CAA CGT TCA GGT GGT TCA

                        420
gly lys gly ser asn lys gly pro leu gly val thr glu glu leu
GGA AAG GGC AGC AAT AAG GGG CCA CTA GGT GTG ACA GAG GAA CTG

    430                                             440
his ile ile ser phe thr val lys tyr thr tyr gln gly leu lys
CAC ATC ATC AGC TTC ACG GTC AAA TAT ACC TAC CAG GGT CTG AAG

                        450
gln glu leu lys thr asp thr leu pro val val ile ile ser asn
CAG GAG CTG AAA ACG GAC ACC CTC CCT GTG GTG ATT ATT TCC AAC

    460                                             470
met asn gln leu ser ile ala trp ala ser val leu trp phe asn
ATG AAC CAG CTC TCA ATT GCC TGG GCT TCA GTT CTC TGG TTC AAT

                        480
leu leu ser pro asn leu gln asn gln gln phe phe ser asn pro
TTG CTC AGC CCA AAC CTT CAG AAC CAG CAG TTC TTC TCC AAC CCC

    490                                             500
pro lys ala pro trp ser leu leu gly pro ala leu ser trp gln
CCC AAG GCC CCC TGG AGC TTG CTG GGC CCT GCT CTC AGT TGG CAG

                        510
phe ser ser tyr val gly arg gly leu asn ser asp gln leu ser
TTC TCC TCC TAT GTT GGC CGA GGC CTC AAC TCA GAC CAG CTG AGC

    520                                             530
met leu arg asn lys leu phe gly gln asn cys arg thr glu asp
ATG CTG AGA AAC AAG CTG TTC GGG CAG AAC TGT AGG ACT GAG GAT

                        540
pro leu leu ser trp ala asp phe thr lys arg glu ser pro pro
CCA TTA TTG TCC TGG GCT GAC TTC ACT AAG CGA GAG AGC CCT CCT

    550                                             560
gly lys leu pro phe trp thr trp leu asp lys ile leu glu leu
GGC AAG TTA CCA TTC TGG ACA TGG CTG GAC AAA ATT CTG GAG TTG

                        570
val his asp his leu lys asp leu trp asn asp gly arg ile met
GTA CAT GAC CAC CTG AAG GAT CTC TGG AAT GAT GGA CGC ATC ATG

    580                                             590
gly phe val ser arg ser gln glu arg arg leu leu lys lys thr
GGC TTT GTG AGT CGG AGC CAG GAG CGC CGG CTG CTG AAG AAG ACC

                        600
met ser gly thr phe leu leu arg phe ser glu ser ser glu gly
ATG TCT GGC ACC TTT CTA CTG CGC TTC AGT GAA TCG TCA GAA GGG
```

Session Name: rb

FIG. 1D

```
    610                                                 620
gly ile thr cys ser trp val glu his gln asp asp asp lys val
GGC ATT ACC TGC TCC TGG GTG GAG CAC CAG GAT GAT GAC AAG GTG

                        630
leu ile tyr ser val gln pro tyr thr lys glu val leu gln ser
CTC ATC TAC TCT GTG CAA CCG TAC ACG AAG GAG GTG CTG CAG TCA

    640                                                 650
leu pro leu thr glu ile ile arg his tyr gln leu leu thr glu
CTC CCG CTG ACT GAA ATC ATC CGC CAT TAC CAG TTG CTC ACT GAG

                        660
glu asn ile pro glu asn pro leu arg phe leu tyr pro arg ile
GAG AAT ATA CCT GAA AAC CCA CTG CGC TTC CTC TAT CCC CGA ATC

    670                                                 680
pro arg asp glu ala phe gly cys tyr tyr gln glu lys val asn
CCC CGG GAT GAA GCT TTT GGG TGC TAC TAC CAG GAG AAA GTT AAT

                        690
leu gln glu arg arg lys tyr leu lys his arg leu ile val val
CTC CAG GAA CGG AGG AAA TAC CTG AAA CAC AGG CTC ATT GTG GTC

    700                                                 710
ser asn arg gln val asp glu leu gln gln pro leu glu leu lys
TCT AAT AGA CAG GTG GAT GAA CTG CAA CAA CCG CTG GAG CTT AAG

                        720
pro glu pro glu leu glu ser leu glu leu glu leu gly leu val
CCA GAG CCA GAG CTG GAG TCA TTA GAG CTG GAA CTA GGG CTG GTG

    730                                                 740
pro glu pro glu leu ser leu asp leu glu pro leu leu lys ala
CCA GAG CCA GAG CTC AGC CTG GAC TTA GAG CCA CTG CTG AAG GCA

                        750
gly leu asp leu gly pro glu leu glu ser val leu glu ser thr
GGG CTG GAT CTG GGG CCA GAG CTA GAG TCT GTG CTG GAG TCC ACT

    760                                                 770
leu glu pro val ile glu pro thr leu cys met val ser gln thr
CTG GAG CCT GTG ATA GAG CCC ACA CTA TGC ATG GTA TCA CAA ACA

                        780
val pro glu pro asp gln gly pro val ser gln pro val pro glu
GTG CCA GAG CCA GAC CAA GGA CCT GTA TCA CAG CCA GTG CCA GAG

    790                                                 800
pro asp leu pro cys asp leu arg his leu asn thr glu pro met
CCA GAT TTG CCC TGT GAT CTG AGA CAT TTG AAC ACT GAG CCA ATG

                        810
glu ile phe arg asn cys val lys ile glu glu ile met pro asn
GAA ATC TTC AGA AAC TGT GTA AAG ATT GAA GAA ATC ATG CCG AAT
```

FIG. 1E

Session Name: rb

```
    820                                                 830
gly asp pro leu leu ala gly gln asn thr val asp glu val tyr
GGT GAC CCA CTG TTG GCT GGC CAG AAC ACC GTG GAT GAG GTT TAC

                        840
val ser arg pro ser his phe tyr thr asp gly pro leu met pro
GTC TCC CGC CCC AGC CAC TTC TAC ACT GAT GGA CCC TTG ATG CCT

    850 851
ser asp phe AM
TCT GAC TTC TAG GAACCACATTTCCTCTGTTCTTTTCATATCTCTTTGCCCTTCCTA

CTCCTCATAGCATGATATTGTTCTCCAAGGATGGGAATCAGGCATGTGTCCCTTCCAAGC

TGTGTTAACTGTTCAAACTCAGGCCTGTGTGACTCCATTGGGGTGAGAGGTGAAAGCATA

ACATGGGTACAGAGGGGACAACAATGAATCAGAACAGATGCTGAGCCATAGGTCTAAATA

GGATCCTGGAGGCTGCCTGCTGTGCTGGGAGGTATAGGGGTCCTGGGGGCAGGCCAGGGC

AGTTGACAGGTACTTGGAGGGCTCAGGGCAGTGGCTTCTTTCCAGTATGGAAGGATTTCA

ACATTTTAATAGTTGGTTAGGCTAAACTGGTGCATACTGGCATTGGCCTTGGTGGGGAGC

ACAGACACAGGATAGGACTCCATTTCTTTCTTCCATTCCTTCATGTCTAGGATAACTTGC

TTTCTTCTTTCCTTTACTCCTGGCTCAAGCCCTGAATTTCTTCTTTTCCTGCAGGGGTTG

AGAGCTTTCTGCCTTAGCCTACCATGTGAAACTCTACCCTGAAGAAAGGGATGGATAGGA

AGTAGACCTCTTTTTCTTACCAGTCTCCTCCCCTACTCTGCCCCCTAAGCTGGCTGTACC

TGTTCCTCCCCCATAAAATGATCCTGCCAATCTAAAAAAAAAA
```

FIG. 2A

```
ATTAAACCTCTCGCCGAGCCCCTCCGCAGACTCTGCGCCGGAAAGTTTCATTTGCTGTATGCCA

TCCTCGAGAGCTGTCTAGGTTAACGTTCGCACTCTGTGTATATAACCTCGACAGTCTTGGCACC

TAACGTGCTGTGCGTAGCTGCTCCTTTGGTTGAATCCCCAGGCCCTTGTTGGGGCACAAGGTGG

     Met Ser Gln Trp Tyr Glu Leu Gln Gln Leu Asp Ser Lys Phe Leu
CAGG ATG TCT CAG TGG TAC GAA CTT CAG CAG CTT GAC TCA AAA TTC CTG

 Glu Gln Val His Gln Leu Tyr Asp Asp Ser Phe Pro Met Glu Ile Arg
 GAG CAG GTT CAC CAG CTT TAT GAT GAC AGT TTT CCC ATG GAA ATC AGA

 Gln Tyr Leu Ala Gln Trp Leu Glu Lys Gln Asp Trp Glu His Ala Ala
 CAG TAC CTG GCA CAG TGG TTA GAA AAG CAA GAC TGG GAG CAC GCT GCC

 Asn Asp Val Ser Phe Ala Thr Ile Arg Phe His Asp Leu Leu Ser Gln
 AAT GAT GTT TCA TTT GCC ACC ATC CGT TTT CAT GAC CTC CTG TCA CAG

 Leu Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn Asn Phe Leu Leu
 CTG GAT GAT CAA TAT AGT CGC TTT TCT TTG GAG AAT AAC TTC TTG CTA

 Gln His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln Asp Asn Phe Gln
 CAG CAT AAC ATA AGG AAA AGC AAG CGT AAT CTT CAG GAT AAT TTT CAG

 Glu Asp Pro Ile Gln Met Ser Met Ile Ile Tyr Ser Cys Leu Lys Glu
 GAA GAC CCA ATC CAG ATG TCT ATG ATC ATT TAC AGC TGT CTG AAG GAA

 Glu Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn Gln Ala Gln Ser
 GAA AGG AAA ATT CTG GAA AAC GCC CAG AGA TTT AAT CAG GCT CAG TCG

 Gly Asn Ile Gln Ser Thr Val Met Leu Asp Lys Gln Lys Glu Leu Asp
 GGG AAT ATT CAG AGC ACA GTG ATG TTA GAC AAA CAG AAA GAG CTT GAC

 Ser Lys Val Arg Asn Val Lys Asp Lys Val Met Cys Ile Glu His Glu
 AGT AAA GTC AGA AAT GTG AAG GAC AAG GTT ATG TGT ATA GAG CAT GAA

 Ile Lys Ser Leu Glu Asp Leu Gln Asp Glu Tyr Asp Phe Lys Cys Lys
 ATC AAG AGC CTG GAA GAT TTA CAA GAT GAA TAT GAC TTC AAA TGC AAA

 Thr Leu Gln Asn Arg Glu His Glu Thr Asn Gly Val Ala Lys Ser Asp
 ACC TTG CAG AAC AGA GAA CAC GAG ACC AAT GGT GTG GCA AAG AGT GAT

 Gln Lys Gln Glu Gln Leu Leu Leu Lys Lys Met Tyr Leu Met Leu Asp
 CAG AAA CAA GAA CAG CTG TTA CTC AAG AAG ATG TAT TTA ATG CTT GAC

Asn Lys Arg Lys Glu Val Val His Lys Ile Ile Glu Leu Leu Asn Val
AAT AAG AGA AAG GAA GTA GTT CAC AAA ATA ATA GAG TTG CTG AAT GTC
```

FIG. 2B

```
Thr Glu Leu Thr Gln Asn Ala Leu Ile Asn Asp Glu Leu Val Glu Trp
ACT GAA CTT ACC CAG AAT GCC CTG ATT AAT GAT GAA CTA GTG GAG TGG

Lys Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro Pro Asn Ala Cys
AAG CGG AGA CAG CAG AGC GCC TGT ATT GGG GGG CCG CCC AAT GCT TGC

Leu Asp Gln Leu Gln Asn Trp Phe Thr Ile Val Ala Glu Ser Leu Gln
TTG GAT CAG CTG CAG AAC TGG TTC ACT ATA GTT GCG GAG AGT CTG CAG

Gln Val Arg Gln Gln Leu Lys Lys Leu Glu Glu Leu Glu Gln Lys Tyr
CAA GTT CGG CAG CAG CTT AAA AAG TTG GAG GAA TTG GAA CAG AAA TAC

Thr Tyr Glu His Asp Pro Ile Thr Lys Asn Lys Gln Val Leu Trp Asp
ACC TAC GAA CAT GAC CCT ATC ACA AAA AAC AAA CAA GTG TTA TGG GAC

Arg Thr Phe Ser Leu Phe Gln Gln Leu Ile Gln Ser Ser Phe Val Val
CGC ACC TTC AGT CTT TTC CAG CAG CTC ATT CAG AGC TCG TTT GTG GTG

Glu Arg Gln Pro Cys Met Pro Thr His Pro Gln Arg Pro Leu Val Leu
GAA AGA CAG CCC TGC ATG CCA ACG CAC CCT CAG AGG CCG CTG GTC TTG

Lys Thr Gly Val Gln Phe Thr Val Lys Leu Arg Leu Leu Val Lys Leu
AAG ACA GGG GTC CAG TTC ACT GTG AAG TTG AGA CTG TTG GTG AAA TTG

Gln Glu Leu Asn Tyr Asn Leu Lys Val Lys Val Leu Phe Asp Lys Asp
CAA GAG CTG AAT TAT AAT TTG AAA GTC AAA GTC TTA TTT GAT AAA GAT

Val Asn Glu Arg Asn Thr Val Lys Gly Phe Arg Lys Phe Asn Ile Leu
GTG AAT GAG AGA AAT ACA GTA AAA GGA TTT AGG AAG TTC AAC ATT TTG

Gly Thr His Thr Lys Val Met Asn Met Glu Glu Ser Thr Asn Gly Ser
GGC ACG CAC ACA AAA GTG ATG AAC ATG GAG GAG TCC ACC AAT GGC AGT

Leu Ala Ala Glu Phe Arg His Leu Gln Leu Lys Glu Gln Lys Asn Ala
CTG GCG GCT GAA TTT CGG CAC CTG CAA TTG AAA GAA CAG AAA AAT GCT

Gly Thr Arg Thr Asn Glu Gly Pro Leu Ile Val Thr Glu Glu Leu His
GGC ACC AGA ACG AAT GAG GGT CCT CTC ATC GTT ACT GAA GAG CTT CAC

Ser Leu Ser Phe Glu Thr Gln Leu Cys Gln Pro Gly Leu Val Ile Asp
TCC CTT AGT TTT GAA ACC CAA TTG TGC CAG CCT GGT TTG GTA ATT GAC

Leu Glu Thr Thr Ser Leu Pro Val Val Val Ile Ser Asn Val Ser Gln
CTC GAG ACG ACC TCT CTG CCC GTT GTG GTG ATC TCC AAC GTC AGC CAG

Leu Pro Ser Gly Trp Ala Ser Ile Leu Trp Tyr Asn Met Leu Val Ala
CTC CCG AGC GGT TGG GCC TCC ATC CTT TGG TAC AAC ATG CTG GTG GCG

Glu Pro Arg Asn Leu Ser Phe Phe Leu Thr Pro Pro Cys Ala Arg Trp
GAA CCC AGG AAT CTG TCC TTC TTC CTG ACT CCA CCA TGT GCA CGA TGG
```

FIG. 2C

```
Ala Gln Leu Ser Glu Val Leu Ser Trp Gln Phe Ser Ser Val Thr Lys
GCT CAG CTT TCA GAA GTG CTG AGT TGG CAG TTT TCT TCT GTC ACC AAA

Arg Gly Leu Asn Val Asp Gln Leu Asn Met Leu Gly Glu Lys Leu Leu
AGA GGT CTC AAT GTG GAC CAG CTG AAC ATG TTG GGA GAG AAG CTT CTT

Gly Pro Asn Ala Ser Pro Asp Gly Leu Ile Pro Trp Thr Arg Phe Cys
GGT CCT AAC GCC AGC CCC GAT GGT CTC ATT CCG TGG ACG AGG TTT TGT

Lys Glu Asn Ile Asn Asp Lys Asn Phe Pro Phe Trp Leu Trp Ile Glu
AAG GAA AAT ATA AAT GAT AAA AAT TTT CCC TTC TGG CTT TGG ATT GAA

Ser Ile Leu Glu Leu Ile Lys Lys His Leu Leu Pro Leu Trp Asn Asp
AGC ATC CTA GAA CTC ATT AAA AAA CAC CTG CTC CCT CTC TGG AAT GAT

Gly Cys Ile Met Gly Phe Ile Ser Lys Glu Arg Glu Arg Ala Leu Leu
GGG TGC ATC ATG GGC TTC ATC AGC AAG GAG CGA GAG CGT GCC CTG TTG

Lys Asp Gln Gln Pro Gly Thr Phe Leu Leu Arg Phe Ser Glu Ser Ser
AAG GAC CAG CAG CCG GGG ACC TTC CTG CTG CGG TTC AGT GAG AGC TCC

Arg Glu Gly Ala Ile Thr Phe Thr Trp Val Glu Arg Ser Gln Asn Gly
CGG GAA GGG GCC ATC ACA TTC ACA TGG GTG GAG CGG TCC CAG AAC GGA

Gly Glu Pro Asp Phe His Ala Val Glu Pro Tyr Thr Lys Lys Glu Leu
GGC GAA CCT GAC TTC CAT GCG GTT GAA CCC TAC ACG AAG AAA GAA CTT

Ser Ala Val Thr Phe Pro Asp Ile Ile Arg Asn Tyr Lys Val Met Ala
TCT GCT GTT ACT TTC CCT GAC ATC ATT CGC AAT TAC AAA GTC ATG GCT

Ala Glu Asn Ile Pro Glu Asn Pro Leu Lys Tyr Leu Tyr Pro Asn Ile
GCT GAG AAT ATT CCT GAG AAT CCC CTG AAG TAT CTG TAT CCA AAT ATT

Asp Lys Asp His Ala Phe Gly Lys Tyr Tyr Ser Arg Pro Lys Glu Ala
GAC AAA GAC CAT GCC TTT GGA AAG TAT TAC TCC AGG CCA AAG GAA GCA

Pro Glu Pro Met Glu Leu Asp Gly Pro Lys Gly Thr Gly Tyr Ile Lys
CCA GAG CCA ATG GAA CTT GAT GGC CCT AAA GGA ACT GGA TAT ATC AAG

Thr Glu Leu Ile Ser Val Ser Glu Val His Pro Ser Arg Leu Gln Thr
ACT GAG TTG ATT TCT GTG TCT GAA GTT CAC CCT TCT AGA CTT CAG ACC

Thr Asp Asn Leu Leu Pro Met Ser Pro Glu Glu Phe Asp Glu Val Ser
ACA GAC AAC CTG CTC CCC ATG TCT CCT GAG GAG TTT GAC GAG GTG TCT

Arg Ile Val Gly Ser Val Glu Phe Asp Ser Met Met Asn Thr Val
CGG ATA GTG GGC TCT GTA GAA TTC GAC AGT ATG ATG AAC ACA GTA TAG

AGCATGAATTTTTTTCATCTTCTCTGGCGACAGTTTTCCTTCTCATCTGTGATTCCCTCCTGCT
```

FIG. 2D

ACTCTGTTCCTTCACATCCTGTGTTTCTAGGGAAATGAAAGAAAGGCCAGCAAATTCGCTGCA

ACCTGTTGATAGCAAGTGAATTTTTCTCTAACTCAGAAACATCAGTTACTCTGAAGGGCATCA

TGCATCTTACTGAAGGTAAAATTGAAAGGCATTCTCTGAAGAGTGGGTTTCACAAGTGAAAAA

CATCCAGATACACCCAAAGTATCAGGACGAGAATGAGGGTCCTTTGGGAAAGGAGAAGTTAAG

CAACATCTAGCAAATGTTATGCATAAAGTCAGTGCCCAACTGTTATAGGTTGTTGGATAAATC

AGTGGTTATTTAGGGAACTGCTTGACGTAGGAACGGTAAATTTCTGTGGGAGAATTCTTACAT

GTTTTCTTTGCTTTAAGTGTAACTGGCAGTTTTCCATTGGTTTACCTGTGAAATAGTTCAAAG

CCAAGTTTATATACAATTATATCAGTCCTCTTTCAAAGGTAGCCATCATGGATCTGGTAGGGG

GAAAATGTGTATTTTATTACATCTTTCACATTGGCTATTTAAAGACAAAGACAAATTCTGTTT

CTTGAGAAGAGAAATTTCCAAATTCACAAGTTGTGTTTGATATCCAAAGCTGAATACATTCTG

CTTTCATCTTGGTCACATACAATTATTTTTACAGTTCTCCCAAGGGAGTTAGGCTATTCACAA

CCACTCATTCAAAAGTTGAAATTAACCATAGATGTAGATAAACTCAGAAATTTAATTCATGTT

TCTTAAATGGGCTACTTTGTCCTTTTTGTTATTAGGGTGGTATTTAGTCTATTAGCCACAAAA

TTGGGAAAGGAGTAGAAAAAGCAGTAACTGACAACTTGAATAATACACCAGAGATAATATGAG

AATCAGATCATTTCAAAACTCATTTCCTATGTAACTGCATTGAGAACTGCATATGTTTCGCTG

ATATATGTGTTTTTCACATTTGCGAATGGTTCCATTCTCTCTCCTGTACTTTTTCCAGACACT

TTTTTGAGTGGATGATGTTTCGTGAAGTATACTGTATTTTTACCTTTTTCCTTCCTTATCACT

GACACAAAAAGTAGATTAAGAGATGGGTTTGACAAGGTTCTTCCCTTTTACATACTGCTGTCT

ATGTGGCTGTATCTTGTTTTTCCACTACTGCTACCACAACTATATTATCATGCAAATGCTGTA

TTCTTCTTTGGTGGAGATAAAGATTTCTTGAGTTTTGTTTTAAAATTAAAGCTAAAGTATCTG

TATTGCATTAAATATAATATCGACACAGTGCTTTCCGTGGCACTGCATACAATCTGAGGCCTC

CTCTCTCAGTTTTTATATAGATGGCGAGAACCTAAGTTTCAGTTGATTTTACAATTGAAATGA

CTAAAAAACAAAGAAGACAACATTAAAAACAATATTGTTTCTA

FIG. 3A

```
ATTAAACCTCTCGCCGAGCCCCTCCGCAGACTCTGCGCCGGAAAGTTTCATTTGCTGTATGCC
ATCCTCGAGAGCTGTCTAGGTTAACGTTCGCACTCTGTGTATATAACCTCGACAGTCTTGGCA
CCTAACGTGCTGTGCGTAGCTGCTCCTTTGGTTGAATCCCCAGGCCCTTGTTGGGGCACAAGG

         Met Ser Gln Trp Tyr Glu Leu Gln Gln Leu Asp Ser Lys Phe
TGGCAGG ATG TCT CAG TGG TAC GAA CTT CAG CAG CTT GAC TCA AAA TTC

Leu Glu Gln Val His Gln Leu Tyr Asp Asp Ser Phe Pro Met Glu Ile
CTG GAG CAG GTT CAC CAG CTT TAT GAT GAC AGT TTT CCC ATG GAA ATC

Arg Gln Tyr Leu Ala Gln Trp Leu Glu Lys Gln Asp Trp Glu His Ala
AGA CAG TAC CTG GCA CAG TGG TTA GAA AAG CAA GAC TGG GAG CAC GCT

Ala Asn Asp Val Ser Phe Ala Thr Ile Arg Phe His Asp Leu Leu Ser
GCC AAT GAT GTT TCA TTT GCC ACC ATC CGT TTT CAT GAC CTC CTG TCA

Gln Leu Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn Asn Phe Leu
CAG CTG GAT GAT CAA TAT AGT CGC TTT TCT TTG GAG AAT AAC TTC TTG

Leu Gln His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln Asp Asn Phe
CTA CAG CAT AAC ATA AGG AAA AGC AAG CGT AAT CTT CAG GAT AAT TTT

Gln Glu Asp Pro Ile Gln Met Ser Met Ile Ile Tyr Ser Cys Leu Lys
CAG GAA GAC CCA ATC CAG ATG TCT ATG ATC ATT TAC AGC TGT CTG AAG

Glu Glu Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn Gln Ala Gln
GAA GAA AGG AAA ATT CTG GAA AAC GCC CAG AGA TTT AAT CAG GCT CAG

Ser Gly Asn Ile Gln Ser Thr Val Met Leu Asp Lys Gln Lys Glu Leu
TCG GGG AAT ATT CAG AGC ACA GTG ATG TTA GAC AAA CAG AAA GAG CTT

Asp Ser Lys Val Arg Asn Val Lys Asp Lys Val Met Cys Ile Glu His
GAC AGT AAA GTC AGA AAT GTG AAG GAC AAG GTT ATG TGT ATA GAG CAT

Glu Ile Lys Ser Leu Glu Asp Leu Gln Asp Glu Tyr Asp Phe Lys Cys
GAA ATC AAG AGC CTG GAA GAT TTA CAA GAT GAA TAT GAC TTC AAA TGC

Lys Thr Leu Gln Asn Arg Glu His Glu Thr Asn Gly Val Ala Lys Ser
AAA ACC TTG CAG AAC AGA GAA CAC GAG ACC AAT GGT GTG GCA AAG AGT

Asp Gln Lys Gln Glu Gln Leu Leu Leu Lys Lys Met Tyr Leu Met Leu
GAT CAG AAA CAA GAA CAG CTG TTA CTC AAG AAG ATG TAT TTA ATG CTT

Asp Asn Lys Arg Lys Glu Val Val His Lys Ile Ile Glu Leu Leu Asn
GAC AAT AAG AGA AAG GAA GTA GTT CAC AAA ATA ATA GAG TTG CTG AAT

Val Thr Glu Leu Thr Gln Asn Ala Leu Ile Asn Asp Glu Leu Val Glu
GTC ACT GAA CTT ACC CAG AAT GCC CTG ATT AAT GAT GAA CTA GTG GAG
```

FIG. 3B

```
Trp Lys Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro Pro Asn Ala
TGG AAG CGG AGA CAG CAG AGC GCC TGT ATT GGG GGG CCG CCC AAT GCT

Cys Leu Asp Gln Leu Gln Asn Trp Phe Thr Ile Val Ala Glu Ser Leu
TGC TTG GAT CAG CTG CAG AAC TGG TTC ACT ATA GTT GCG GAG AGT CTG

Gln Gln Val Arg Gln Gln Leu Lys Lys Leu Glu Glu Leu Glu Gln Lys
CAG CAA GTT CGG CAG CAG CTT AAA AAG TTG GAG GAA TTG GAA CAG AAA

Tyr Thr Tyr Glu His Asp Pro Ile Thr Lys Asn Lys Gln Val Leu Trp
TAC ACC TAC GAA CAT GAC CCT ATC ACA AAA AAC AAA CAA GTG TTA TGG

Asp Arg Thr Phe Ser Leu Phe Gln Gln Leu Ile Gln Ser Ser Phe Val
GAC CGC ACC TTC AGT CTT TTC CAG CAG CTC ATT CAG AGC TCG TTT GTG

Val Glu Arg Gln Pro Cys Met Pro Thr His Pro Gln Arg Pro Leu Val
GTG GAA AGA CAG CCC TGC ATG CCA ACG CAC CCT CAG AGG CCG CTG GTC

Leu Lys Thr Gly Val Gln Phe Thr Val Lys Leu Arg Leu Leu Val Lys
TTG AAG ACA GGG GTC CAG TTC ACT GTG AAG TTG AGA CTG TTG GTG AAA

Leu Gln Glu Leu Asn Tyr Asn Leu Lys Val Lys Val Leu Phe Asp Lys
TTG CAA GAG CTG AAT TAT AAT TTG AAA GTC AAA GTC TTA TTT GAT AAA

Asp Val Asn Glu Arg Asn Thr Val Lys Gly Phe Arg Lys Phe Asn Ile
GAT GTG AAT GAG AGA AAT ACA GTA AAA GGA TTT AGG AAG TTC AAC ATT

Leu Gly Thr His Thr Lys Val Met Asn Met Glu Glu Ser Thr Asn Gly
TTG GGC ACG CAC ACA AAA GTG ATG AAC ATG GAG GAG TCC ACC AAT GGC

Ser Leu Ala Ala Glu Phe Arg His Leu Gln Leu Lys Glu Gln Lys Asn
AGT CTG GCG GCT GAA TTT CGG CAC CTG CAA TTG AAA GAA CAG AAA AAT

Ala Gly Thr Arg Thr Asn Glu Gly Pro Leu Ile Val Thr Glu Glu Leu
GCT GGC ACC AGA ACG AAT GAG GGT CCT CTC ATC GTT ACT GAA GAG CTT

His Ser Leu Ser Phe Glu Thr Gln Leu Cys Gln Pro Gly Leu Val Ile
CAC TCC CTT AGT TTT GAA ACC CAA TTG TGC CAG CCT GGT TTG GTA ATT

Asp Leu Glu Thr Thr Ser Leu Pro Val Val Val Ile Ser Asn Val Ser
GAC CTC GAG ACG ACC TCT CTG CCC GTT GTG GTG ATC TCC AAC GTC AGC

Gln Leu Pro Ser Gly Trp Ala Ser Ile Leu Trp Tyr Asn Met Leu Val
CAG CTC CCG AGC GGT TGG GCC TCC ATC CTT TGG TAC AAC ATG CTG GTG

Ala Glu Pro Arg Asn Leu Ser Phe Phe Leu Thr Pro Pro Cys Ala Arg
GCG GAA CCC AGG AAT CTG TCC TTC TTC CTG ACT CCA CCA TGT GCA CGA

Trp Ala Gln Leu Ser Glu Val Leu Ser Trp Gln Phe Ser Ser Val Thr
TGG GCT CAG CTT TCA GAA GTG CTG AGT TGG CAG TTT TCT TCT GTC ACC
```

FIG. 3C

```
Lys Arg Gly Leu Asn Val Asp Gln Leu Asn Met Leu Gly Glu Lys Leu
AAA AGA GGT CTC AAT GTG GAC CAG CTG AAC ATG TTG GGA GAG AAG CTT

Leu Gly Pro Asn Ala Ser Pro Asp Gly Leu Ile Pro Trp Thr Arg Phe
CTT GGT CCT AAC GCC AGC CCC GAT GGT CTC ATT CCG TGG ACG AGG TTT

Cys Lys Glu Asn Ile Asn Asp Lys Asn Phe Pro Phe Trp Leu Trp Ile
TGT AAG GAA AAT ATA AAT GAT AAA AAT TTT CCC TTC TGG CTT TGG ATT

Glu Ser Ile Leu Glu Leu Ile Lys Lys His Leu Leu Pro Leu Trp Asn
GAA AGC ATC CTA GAA CTC ATT AAA AAA CAC CTG CTC CCT CTC TGG AAT

Asp Gly Cys Ile Met Gly Phe Ile Ser Lys Glu Arg Glu Arg Ala Leu
GAT GGG TGC ATC ATG GGC TTC ATC AGC AAG GAG CGA GAG CGT GCC CTG

Leu Lys Asp Gln Gln Pro Gly Thr Phe Leu Leu Arg Phe Ser Glu Ser
TTG AAG GAC CAG CAG CCG GGG ACC TTC CTG CTG CGG TTC AGT GAG AGC

Ser Arg Glu Gly Ala Ile Thr Phe Thr Trp Val Glu Arg Ser Gln Asn
TCC CGG GAA GGG GCC ATC ACA TTC ACA TGG GTG GAG CGG TCC CAG AAC

Gly Gly Glu Pro Asp Phe His Ala Val Glu Pro Tyr Thr Lys Lys Glu
GGA GGC GAA CCT GAC TTC CAT GCG GTT GAA CCC TAC ACG AAG AAA GAA

Leu Ser Ala Val Thr Phe Pro Asp Ile Ile Arg Asn Tyr Lys Val Met
CTT TCT GCT GTT ACT TTC CCT GAC ATC ATT CGC AAT TAC AAA GTC ATG

Ala Ala Glu Asn Ile Pro Glu Asn Pro Leu Lys Tyr Leu Tyr Pro Asn
GCT GCT GAG AAT ATT CCT GAG AAT CCC CTG AAG TAT CTG TAT CCA AAT

Ile Asp Lys Asp His Ala Phe Gly Lys Tyr Tyr Ser Arg Pro Lys Glu
ATT GAC AAA GAC CAT GCC TTT GGA AAG TAT TAC TCC AGG CCA AAG GAA

Ala Pro Glu Pro Met Glu Leu Asp Gly Pro Lys Gly Thr Gly Tyr Ile
GCA CCA GAG CCA ATG GAA CTT GAT GGC CCT AAA GGA ACT GGA TAT ATC

Lys Thr Glu Leu Ile Ser Val Ser Glu Val

AAG ACT GAG TTG ATT TCT GTG TCT GAA GTG TAAGTGAACACAGAAGAGTGACA

TGTTTACAAACCTCAAGCCAGCCTTGCTCCTGGCTGGGGCCTGTTGAAGATGCTTGTATTTTA

CTTTTCCATTGTAATTGCTATCGCCATCACAGCTGAACTTGTTGAGATCCCCGTGTTACTGCC

TATCAGCATTTTACTACTTTAAAAAAAAAAAAAAAAGCCAAAAACCAAATTTGTATTTAAGGT

ATATAAATTTTCCCAAAACTGATACCCTTTGAAAAAGTATAAATAAAATGAGCAAAAGTTGAA
```

FIG. 4

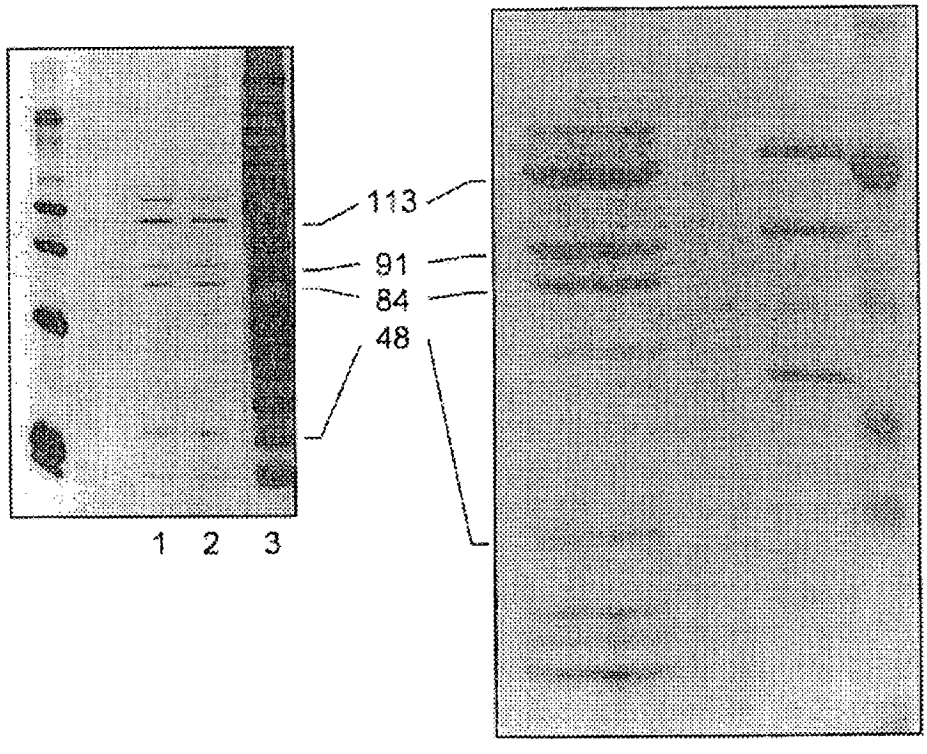


FIG. 5A

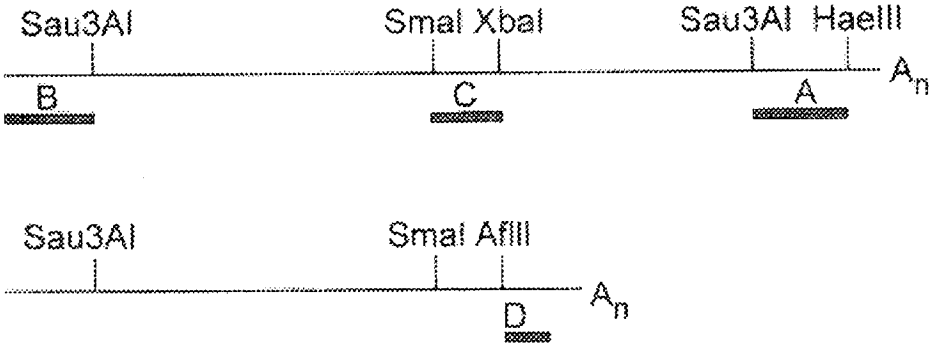

Sau3AI
SmaI XbaI
Sau3AI HaeIII
A_n
B
C
A
Sau3AI
SmaI AflII
A_n
D

FIG. 5B

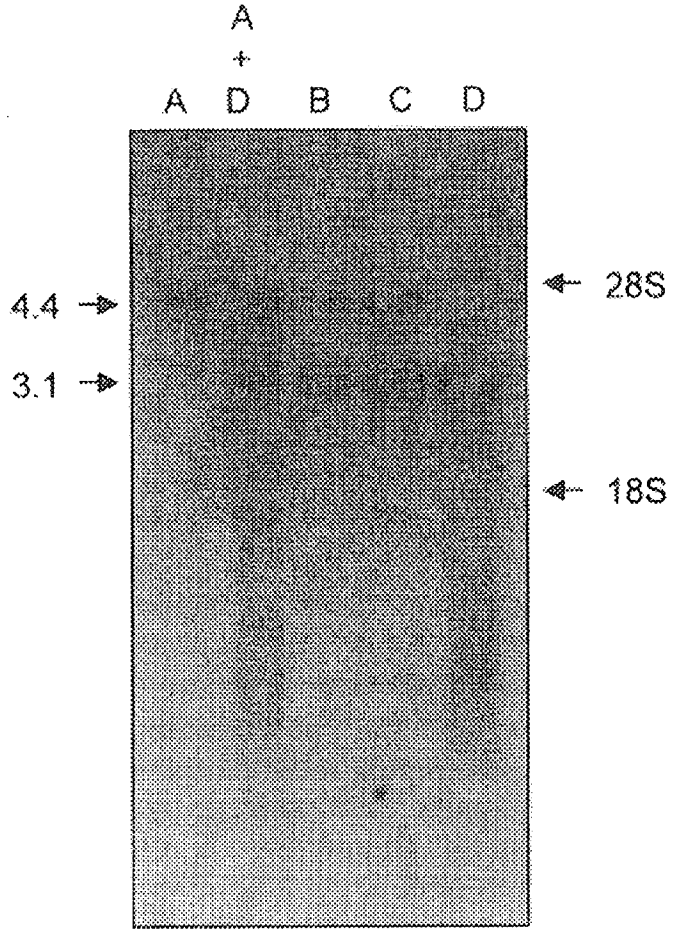

A
+
A D B C D
4.4
3.1
28S
18S

FIG. 6

```
  1 MSQWYELQQLDSKFLEQVHQLYDDSFPMEIRQYLAQWLEKQDWEHAANDV
 51 SFATIRFHDLLSQLDDQYSRFSLENNFLLQHNIRKSKRNLQDNFQEDPIQ
101 MSMIIYSCLKEERKILENAQRFNQAQSGNIQSTVMLDKQKELDSKVRNVK
151 DKVMCIEHEIKSLEDLQDEYDFKCKTLQNREHETNGVAKSDQKQEQLLLK
201 KMYLMLDNKRKEVVHKIIELLNVTELTQNALINDELVEWKRRQQSACIGG
251 PPNACLDQLQQVRQQLKKLEELEQKYTYEHDPITKNKQVLWDRTFSLFQQ
301 LIQSSFVVERQPCMPTHPQRPLVLKTGVQFTVKLRLLVKLQELNYNLKVK
351 VLFDKDVNERNTVKGFRKFNILGTHEKVMNMEESTNGSLAAEFRHLQLKE
401 QKNAGTRTNEGPLIVTEELHSLSFETQLCQPGLVIDLETTSLPVVVISNV
451 SQLPSGWASILWYNMLVAEPRNLSFFLTPPCARWAQLSEVLSWQFSSVTK
        127
501 RGLNVDQLNMLGEKLLGPNASPDGLIPWTRFCKENINDKNFPFWLWIESI
          119
551 LELIKKHLLPLWNDGCIMGFISKERERALLKDQQPGTFLLRFSESSREGA
601 ITFTWVERSQNGGEPDFHAVEPYTKKELSAVTFPDIIRNYKVMAAENIPE
                                                  113a
651 NPLKYLYPNIDKDHAFGKYYSRPKEAPEPMELDGPKGTGYIKTELISVSE
          113b
701 VHPSRLQTTDNLLPMSPEEPDEVSRIVGSVEFDSMMNTV
    ↑
    last amino acid of 84 kd
```

FIG. 7A

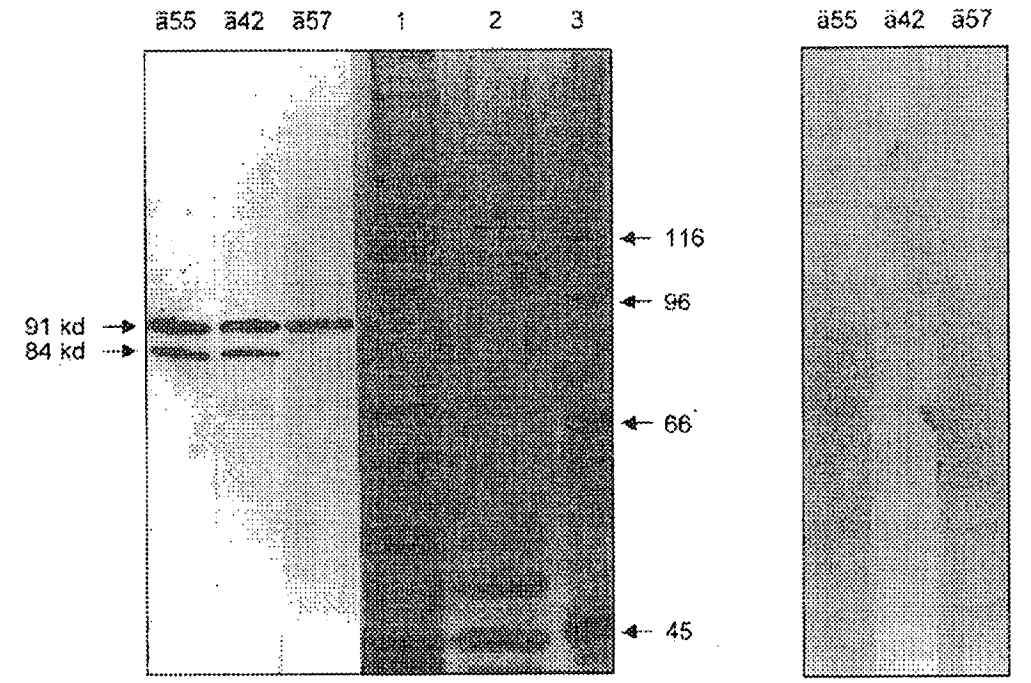


FIG. 7B

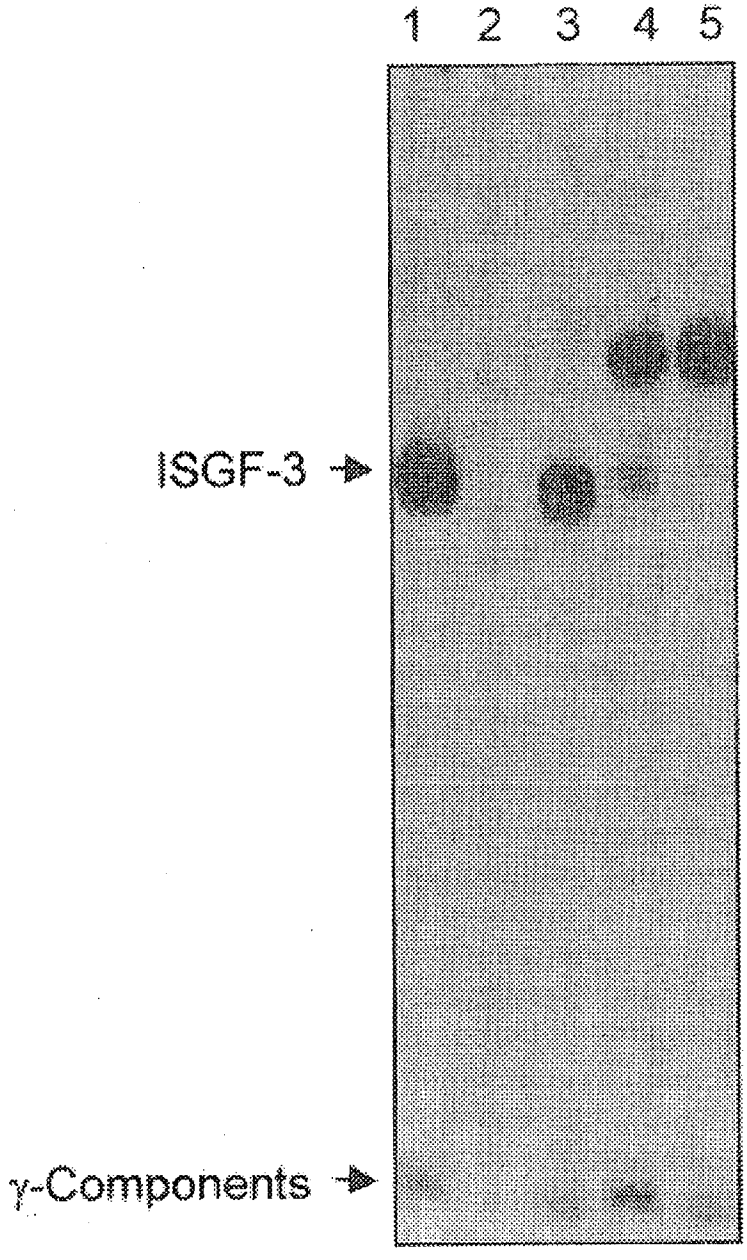


FIG. 8a

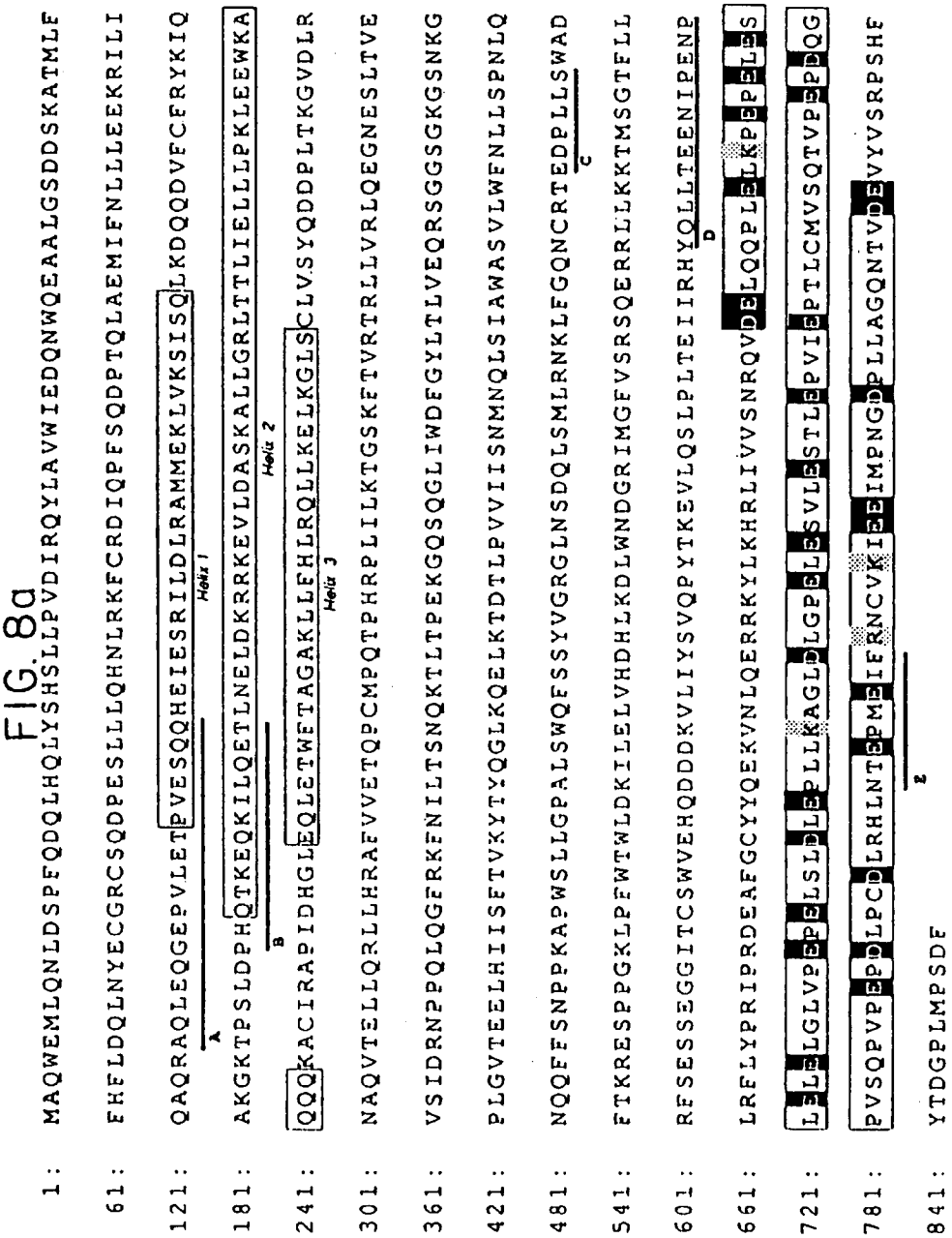

1: MAQWEMLQNLDSPFQDQLHQLYSHSLLPVDIRQYLAVWIEDQNWQEAALGSDDSKATMLF
61: FHFLDQLNYECGRCSQDPESLLLQHNLRKFCRDIQPFSQDPTQLAEMIFNLLLEEKRILI
121: QAQRAQLEQGEPVLETPVESQQHEIESRILDLRAMMEKLVKSISQLKDQQDVFCFRYKIQ
181: AKGKTPSLDPHQTKEQKILQETLNELDKRRKEVLDASKALLGRLTTLIELLLPKLEEWKA
241: QQQKACIRAPIDHGLLEQLETWFTAGAKLLFHLRQLLKELKGLSCLVSYQDDPLTKGVDLR
301: NAQVTELLQRLLHRAFVVETQPCMPQTPHRPLILKTGSKFTVRTRLLVRLQEGNESLTVE
361: VSIDRNPPQLQGFRKFNILTSNQKTLTPEKGQSQGLIWDFGYLTLVEQRSGGSGKGSNKG
421: PLGVTEELHIISFTVKYTYQGLKQELKTDTLPVVIISNMNQLSIAWASVLWFNLLSPNLQ
481: NQQFFSNPPKAPWSLLGPALSWQFSSYVGRGLNSDQLSMLRNKLFGQNCRTEDPLLSWAD
541: FTKRESPPGKLPFWTWLDKILELVHDHLKDLWNDGRIMGFVSRSQERRLLKKTMSGTFLL
601: RFSESSEGGITCSWVEHQDDDKVLIYSVQPYTKEVLQSLPLTEIIRHYQLLTEENIPENP
661: LRFLYPRIPRDEAFGCYYQEKVNLQERRKYLKHRLIVVSNRQVDELQQPLELKPPELES
721: LLELGLVPEPELSLDLEPLLKAGLDLGPELESVLESTLEPVIEPTLCMVSQTVPEPDQG
781: PVSQPVPEPDLPCDLRHLNTEPMEIFRNCVKIEEIMPNGDPLLAGQNTVDEVYVSRPSHF
841: YTDGPLMPSDF

FIG. 8b

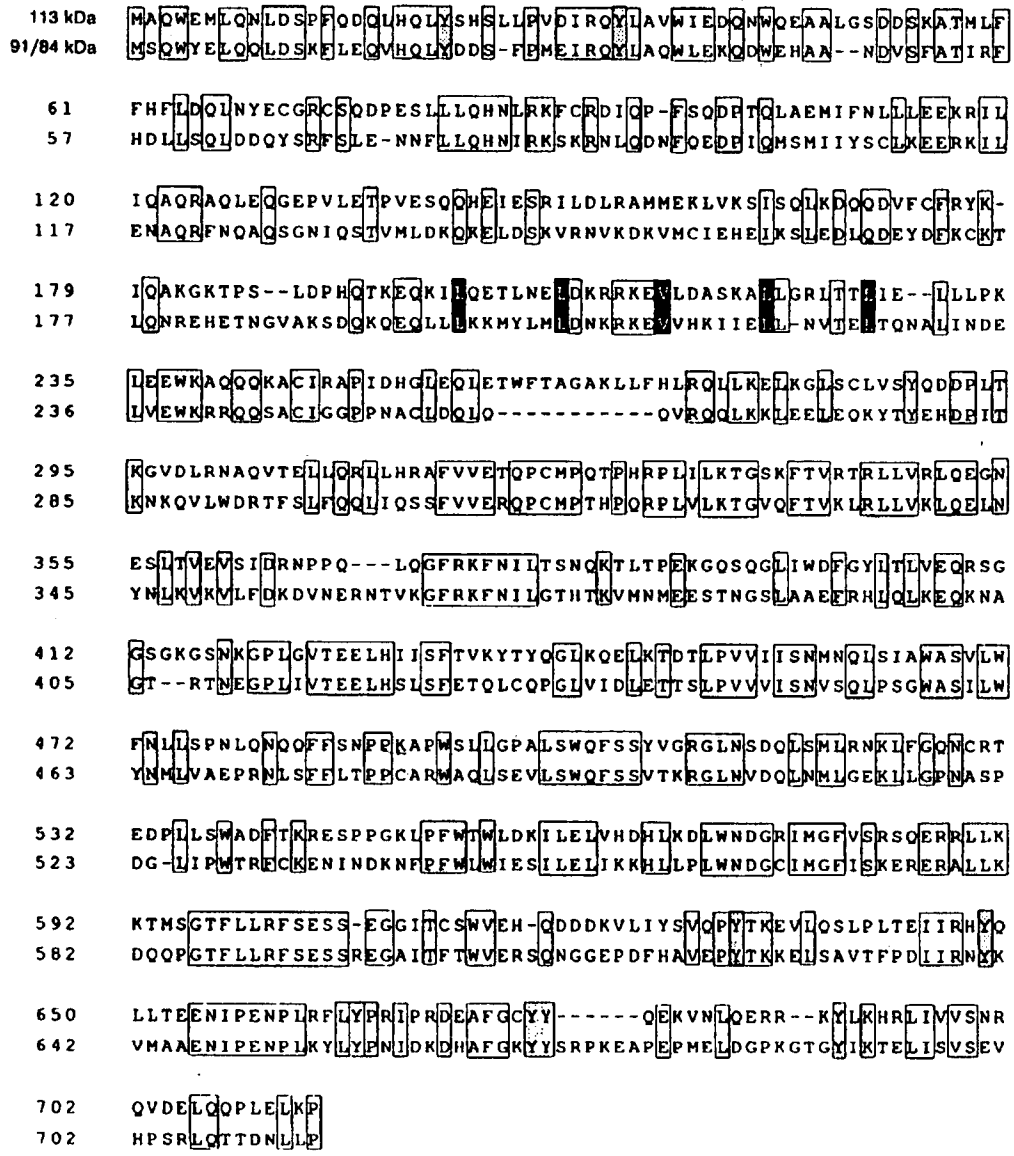

113 kDa
91/84 kDa
61
57
120
117
179
177
235
236
295
285
355
345
412
405
472
463
532
523
592
582
650
642
702
702

FIG. 9A

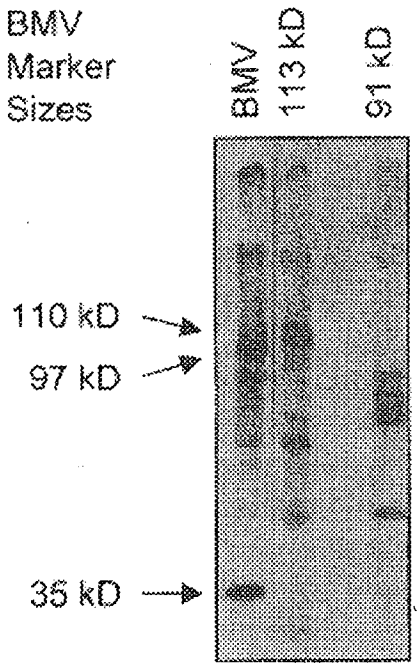


FIG. 9B

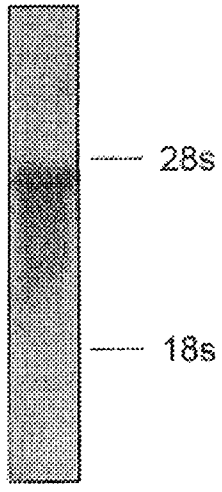


FIG. 10A

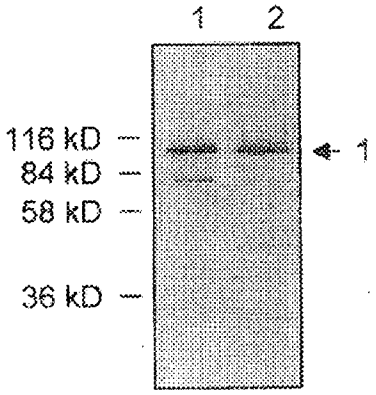


FIG. 10B

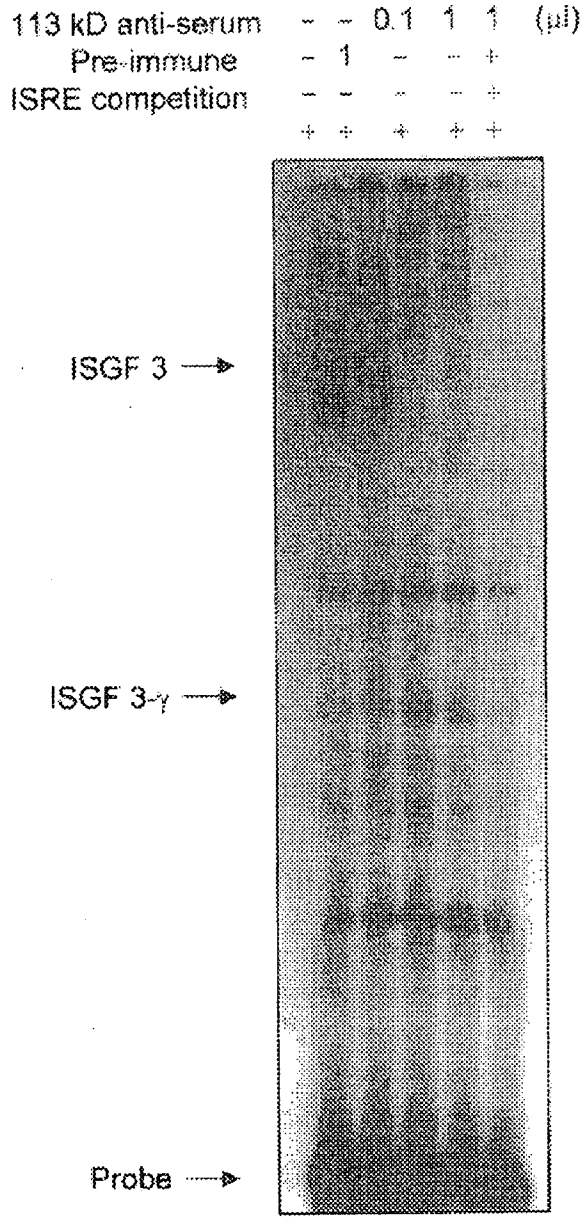


FIG. 11

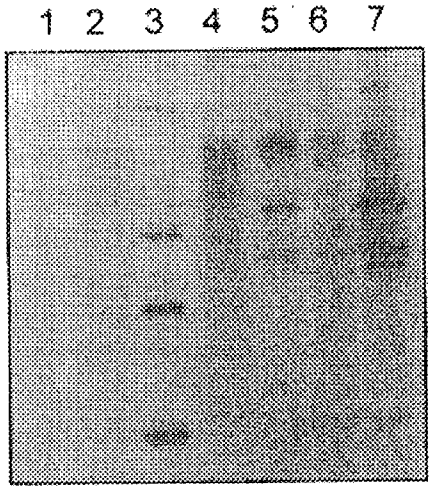


FIG. 12

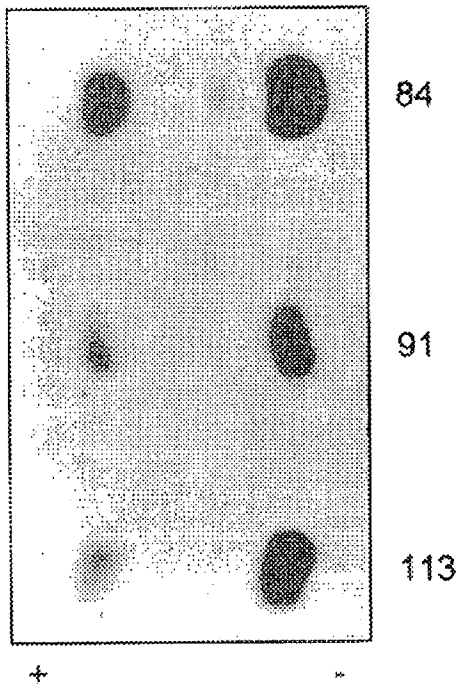


FIG. 13A

Mouse 91kD (protein)

Amino acid sequence (deduced)

1 MSQWFELQQL DSKFLEQVHQ LYDDSFPMEI RQYLAQWLEK QDWEHAAYDV

51 SFATIRFHDL LSQLDDQYSR FSLENNFLLQ HNIRKSKRNL QDNFQEDPVQ

101 MSMIIYNCLK EERKILENAQ RFNQAQEGNI QNTVMLDKQK ELDSKVRNVK

151 DQVMCIEQEI KTLEELQDEY DFKCKTSQNR EGEANGVAKS DQKQEQLLLH

201 KMFLMLDNKR KEIIHKIREL LNSIELTQNT LINDELVEWK RRQQSACIGG

251 PPNACLDQLQ TWFTIVAETL QQIRQQLKKL EELEQKFTYE PDPITKNKQV

301 LSDRTFLLFQ QLIQSSFVVE RQPCMPTHPQ RPLVLKTGVQ FTVKSRLLVK

351 LQESNLLTKV KCHFDKDVNE KNTVKGFRKF NILGTHTKVM NMEESTNGSL

401 AAELRHLQLK EQKNAGNRTN EGPLIVTEEL HSLSFETQLC QPGLVIDLET

451 TSLPVVVISN VSQLPSGWAS ILWYNMLVTE PRNLSFFLNP PCAWWSQLSE

501 VLSWQFSSVT KRGLNADQLS MLGEKLLGPN AGPDGLIPWT RFCKENINDK

551 NFSFWPWIDT ILELIKNDLL CLWNDGCIMG FISKERERAL LKDQQPGTFL

601 LRFSESSREG AITFTWVERS QNGGEPDFHA VEPYTKKELS AVTFPDIIRN

651 YKVMAAENIP ENPLKYLYPN IDKDHAFGKY YSRPKEAPEP MELDDPKRTG

701 YIKTELISVS EVHPSRLQTT DNLLPMSPEE FDEMSRIVGP EFDSMMSTV

FIG. 13B

Mouse 91kD(protein)DNA sequence

```
   1  caggatgtca cagtggttcg agcttcagca gctggactcc aagttcctgg
  51  agcaggtcca ccagctgtac gatgacagtt tccccatgga aatcagacag
 101  tacctggccc agtggctgga aaagcaagac tgggagcacg ctgcctatga
 151  tgtctcgttt gcgaccatcc gcttccatga cctcctctca cagctggacg
 201  accagtacag ccgcttttct ctggagaata atttcttgtt gcagcacaac
 251  atacggaaaa gcaagcgtaa tctccaggat aacttccaag aagatcccgt
 301  acagatgtcc atgatcatct acaactgtct gaaggaagaa aggaagattt
 351  tggaaaatgc ccaaagattt aatcaggccc aggagggaaa tattcagaac
 401  actgtgatgt tagataaaca gaaggagctg gacagtaaag tcagaaatgt
 451  gaaggatcaa gtcatgtgca tagagcagga aatcaagacc ctagaagaat
 501  tacaagatga atatgacttt aaatgcaaaa cctctcagaa cagagaaggt
 551  gaagccaatg gtgtggcgaa gagcgaccaa aaacaggaac agctgctgct
 601  ccacaagatg tttttaatgc ttgacaataa gagaaaggag ataattcaca
```

FIG. 13C

```
 651  aaatcagaga gttgctgaat tccatcgagc tcactcagaa cactctgatt
 701  aatgacgagc tcgtggagtg gaagcgaagg cagcagagcg cctgcatcgg
 751  gggaccgccc aacgcctgcc tggatcagct gcaaacgtgg ttcaccattg
 801  ttgcagagac cctgcagcag atccgtcagc agcttaaaaa gctggaggag
 851  ttggaacaga aattcaccta tgagcccgac cctattacaa aaaacaagca
 901  ggtgttgtca gatcgaacct tcctcctctt ccagcagctc attcagagct
 951  ccttcgtggt agaacgacag ccgtgcatgc ccactcaccc gcagaggccc
1001  ctggtcttga agactggggt acagttcact gtcaagtcga gactgttggt
1051  gaaattgcaa gagtcgaatc tattaacgaa agtgaaatgt cactttgaca
1101  aagatgtgaa cgagaaaaac acagttaaag gatttcggaa gttcaacatc
1151  ttgggtacgc acacaaaagt gatgaacatg gaagaatcca ccaacggaag
1201  tctggcagct gagctccgac acctgcaact gaaggaacag aaaaacgctg
1251  ggaacagaac taatgagggg cctctcattg tcaccgaaga acttcactct
1301  cttagctttg aaacccagtt gtgccagcca ggcttggtga ttgacctgga
1351  gaccacctct cttcctgtcg tggtgatctc caacgtcagc cagctcccca
```

FIG. 13D

```
1401  gtggctgggc gtctatcctg tggtacaaca tgctggtgac agagcccagg
1451  aatctctcct tcttcctgaa ccccccgtgc gcgtggtggt cccagctctc
1501  agaggtgttg agttggcagt tttcatcagt caccaagaga ggtctgaacg
1551  cagaccagct gagcatgctg ggagagaagc tgctgggccc taatgctggc
1601  cctgatggtc ttattccatg gacaaggttt tgtaaggaaa atattaatga
1651  taaaaatttc tccttctggc cttggattga caccatccta gagctcatta
1701  agaacgacct gctgtgcctc tggaatgatg ggtgcattat gggcttcatc
1751  agcaaggagc gagaacgcgc tctgctcaag gaccagcagc cagggacgtt
1801  cctgcttaga ttcagtgaga gctcccggga aggggccatc acattcacat
1851  gggtggaacg gtcccagaac ggaggtgaac ctgacttcca tgccgtggag
1901  ccctacacga aaaaagaact ttcagctgtt actttcccag atattattcg
1951  caactacaaa gtcatggctg ccgagaacat accagagaat cccctgaagt
2001  atctgtaccc caatattgac aaagaccacg cctttgggaa gtattattcc
2051  agaccaaagg aagcaccaga accgatggag cttgacgacc ctaagcgaac
2101  tggatacatc aagactgagt tgatttctgt gtctgaagtc cacccttcta
2151  gacttcagac cacagacaac ctgcttccca tgtctccaga ggagtttgat
2201  gagatgtccc ggatagtggg ccccgaattt gacagtatga tgagcacagt
2251  ataaacacga atttctctct ggcgaca
```

FIG. 14A

13sf1 (protein)

Amino acid sequence of 13sf1

```
  1  MSQWNQVQQL EIKFLEQVDQ FYDDNFPMEI RHLLAQWIET QDWEVASNNE

 51  TMATILLQNL LIQLDEQLGR VSKEKNLLLI HNLKRIRKVL QGKFHGNPMH

101  VAVVISNCLR EERRILAAAN MPIQGPLEKS LQSSSVSERQ RNVEHKVSAI

151  KNSVQMTEQD TKYLEDLQDE FDYRYKTIQT MDQGDKNSIL VNQEVLTLLQ

201  EMLNSLDFKR KEALSKMTQI VNETDLLMNS MLLEELQDWK KRHIRIACIGG

251  PLHNGLDQLQ NCFTLLAESL FQLRQQLEKL QEQSTKMTYE GDPIPAQRAH

301  LLERATFLIY NLFKNSFVVE RHACMPTHPQ RPMVLKTLIQ FTVKLRLLIK

351  LPELNYQVKV KASIDKNVST LSNRRFVLCG THVKAMSSEE SSNGSLSVEL

401  DIATQGDEVQ YWSKGNEGCH MVTEELHSIT FETQICLYGL TINLETSSLP

451  VVMISNVSQL PNAWASIIWY NVSTNDSQNL VFFNNPPSVT LGQLLEVMSW

501  QFSSYVGRGL NSEQLNMLAE KLTVQSNYND GHLTWAKFCK EHLPGKTFTF

551  WTWLEAILDL IKKHILPLWI DGYIMGFVSK EKERLLLKDK MPGTFLLRFS

601  ESHLGGITFT WVDQSENGEV RFHSVEPYNK GRLSALAFAD ILRDYKVIMA

651  ENIPENPLKY LYPDIPKDKA FGKHYSSQPC EVSRPTERGD KGYVPSVFIP

701  ISTIRSDSTE PQSPSDLLPM SPSAYAVLRE NLSPTTIETA MNSPYSAE
```

FIG. 14B

13sf1 (DNA)

DNA sequence of 13sf1

```
   1  tgccactacc tggacggaga gagagagagc agcatgtctc agtggaatca

  51  agtccaacaa ttagaaatca agtttttgga gcaagtagat cagttctatg

 101  atgacaactt tcctatggaa atccggcatc tgctagctca gtggattgag

 151  actcaagact gggaagtagc ttctaacaat gaaactatgg caacaattct

 201  gcttcaaaac ttactaatac aattggatga acagttgggg cgggtttcca

 251  aagaaaaaaa tctgctattg attcacaatc taaagagaat tagaaaagtt

 301  cttcagggca agtttcatgg aaatccaatg catgtagctg tggtaatttc

 351  aaattgctta agggaagaga ggagaatatt ggctgcagcc aacatgccta

 401  tccagggacc tctggagaaa tccttacaga gttcttcagt ttctgaaaga

 451  caaaggaatg tggaacacaa agtgtctgcc attaaaaaca gtgtgcagat

 501  gacagaacaa gataccaaat acttagaaga cctgcaagat gagtttgact

 551  acaggtataa aacaattcag acaatggatc agggtgacaa aaacagtatc

 601  ctggtgaacc aggaagtttt gacactgctg caagaaatgc ttaatagtct

 651  ggacttcaag agaaaggaag cactcagtaa gatgacgcag atagtgaacg

 701  agacagacct gctcatgaac agcatgcttc tagaagagct gcaggactgg

 751  aaaaagcggc acaggattgc ctgcattggt ggcccgctcc acaatgggct

 801  ggaccagctt cagaactgct ttaccctact ggcagagagt cttttccaac

 851  tcagacagca actggagaaa ctacaggagc aatctactaa aatgacctat
```

FIG. 14C

13sf1 (DNA)

901 gaaggggatc ccatccctgc tcaaagagca cacctcctgg aaagagctac

951 cttcctgatc tacaaccttt tcaagaactc atttgtggtc gagcgacacg

1001 catgcatgcc aacgcaccct cagaggccga tggtacttaa aacccтcatt

1051 cagttcactg taaaactgag attactaata aaattgccgg aactaaacta

1101 tcaggtgaaa gtaaaggcgt ccattgacaa gaatgtttca actctaagca

1151 atagaagatt tgtgctttgt ggaactcacg tcaaagctat gtccagtgag

1201 gaatcttcca atgggagcct ctcagtggag ttagacattg caacccaagg

1251 agatgaagtg cagtactgga gtaaaggaaa cgagggctgc cacatggtga

1301 cagaggagtt gcattccata acctttgaga cccagatctg cctctatggc

1351 ctcaccatta acctagagac cagctcatta cctgtcgtga tgatttctaa

1401 tgtcagccaa ctacctaatg catgggcatc catcatttgg tacaatgtat

1451 caactaacga ctcccagaac ttggttttct ttaataaccc tccatctgtc

1501 actttgggcc aactcctgga agtgatgagc tggcaatttt catcctatgt

1551 cggtcgtggc cttaattcag agcagctcaa catgctggca gagaagctca

1601 cagttcagtc taactacaat gatggtcacc tcacctgggc caagttctgc

1651 aaggaacatt tgcctggcaa aacatttacc ttctggactt ggcttgaagc

1701 aatattggac ctaattaaaa aacatattct tcccctctgg attgatgggt

1751 acatcatggg atttgttagt aaagagaagg aacggcttct gctcaaagat

1801 aaaatgcctg ggacattttt gttaagattc agtgagagcc atcttggagg

FIG. 14D

13sf1 (DNA)

1851 gataaccttc acctgggtgg accaatctga aaatggagaa gtgagattcc

1901 actctgtaga accctacaac aaagggagac tgtcggctct ggccttcgct

1951 gacatcctgc gagactacaa ggttatcatg gctgaaaaca tcccctgaaaa

2001 ccctctgaag tacctctacc ctgacattcc caaagacaaa gcctttggca

2051 aacactacag ctcccagccg tgcgaagtct caagaccaac cgaacgggga

2101 gacaagggtt acgtcccctc tgtttttatc cccatttcaa caatccgaag

2151 cgattccacg gagccacaat ctccttcaga ccttctcccc atgtctccaa

2201 gtgcatatgc tgtgctgaga gaaaacctga gcccaacgac aattgaaact

2251 gcaatgaatt ccccatattc tgctgaatga cggtgcaaac ggacacttta

2301 aagaaggaag cagatgaaac tggagagtgt tctttaccat agatcacaat

2351 ttatttcttc ggctttgtaa atacc

FIG. 15A

19sf6(DNA)

Amino acid sequence of 19sf6

```
  1  MAQWNQLQQL DTRYLKQLHQ LYSDTFPMEL RQFLAPWIES QDWAYAASKE

 51  SHATLVFHNL LGEIDQQYSR FLQESNVLYQ HNLRRIKQFL QSRYLEKPME

101  IARIVARCLW EESRLLQTAA TAAQQGGQAN HPTAAVVTEK QQMLEQHLQD

151  VRKRVQDLEQ KMKVVENLQD DFDFNYKTLK SQGDMQDLNG NNQSVTRQKM

201  QQLEQMLTAL DQMRRSIVSE LAGLLSAMEY VQKTLTDEEL ADWKRRPEIA

251  CIGGPPNICL DRLENWITSL AESQLQTRQQ IKKLEELQQK VSYKGDPIVQ

301  HRPMLEERIV ELFRNLMKSA FVVERQPCMP MHPDRPLVIK TGVQFTTKVR

351  LLVKFPELNY QLKIKVCIDK DSGDVAALRG SRKFNILGTN TKVMNMEESN

401  NGSLSAEFKH LTLREQRCGN GGRANCDASL IVTEELHLIT FETEVYHQGL

451  KIDLETHSLP VVVISNICQM PNAWASILWY NMLTNNPKNV NFFTKPPIGT

501  WDQVAEVLSW QFSSTTKRGL SIEQLTTLAE KLLGPGVNYS GCQITWAKFC

551  KENMAGKGFS FWVWLDNIID LVKKYILALW NEGYIMGFIS KERERAILST

601  KPPGTFLLRF SESSKEGGVT FTWVEKDISG KTQIQSVEPY TKQQLNNMSF

651  AEIIMGYKIM DATNILVSPL VYLYPDIPKE EAFGKYCRPE SQEHPEADPG

701  SAAPYLKTKF ICVTPTTCSN TIDLPMSPRT LDSLMQFGNN GEGAEPSAGG

751  QFESLTFDMD LTSECATSPM
```

FIG. 15B

19sf6 (DNA)

Amino acid sequence of 19sf6

```
   1  gccgcgacca gccaggccgg ccagtcgggc tcagcccgga gacagtcgag

  51  acccctgact gcagcaggat ggctcagtgg aaccagctgc agcagctgga

 101  cacacgctac ctgaagcagc tgcaccagct gtacagcgac acgttcccca

 151  tggagctgcg gcagttcctg gcaccttgga ttgagagtca agactgggca

 201  tatgcagcca gcaaagagtc acatgccacg ttggtgtttc ataatctctt

 251  gggtgaaatt gaccagcaat atagccgatt cctgcaagag tccaatgtcc

 301  tctatcagca caaccttcga agaatcaagc agtttctgca gagcaggtat

 351  cttgagaagc caatggaaat tgcccggatc gtggcccgat gcctgtggga

 401  agagtctcgc ctcctccaga cggcagccac ggcagcccag caaggggcc

 451  aggccaacca cccaacagcc gccgtagtga cagagaagca gcagatgttg

 501  gagcagcatc ttcaggatgt ccggaagcga gtgcaggatc tagaacagaa

 551  aatgaaggtg gtggagaacc tccaggacga ctttgatttc aactacaaaa

 601  ccctcaagag ccaaggagac atgcaggatc tgaatggaaa caaccagtct

 651  gtgaccagac agaagatgca gcagctggaa cagatgctca cagccctgga

 701  ccagatgcgg agaagcattg tgagtgagct ggcggggctc ttgtcagcaa

 751  tggagtacgt gcagaagaca ctgactgatg aagagctggc tgactggaag

 801  aggcggccag agatcgcgtg catcggaggc cctcccaaca tctgcctgga

 851  ccgtctggaa aactggataa cttcattagc agaatctcaa cttcagaccc
```

FIG. 15C

19sf6 (DNA)

```
 901  gccaacaaat taagaaactg gaggagctgc agcagaaagt gtcctacaag

 951  ggcgacccta tcgtgcagca ccggcccatg ctggaggaga ggatcgtgga

1001  gctgttcaga aacttaatga agagtgcctt cgtggtggag cggcagccct

1051  gcatgcccat gcacccggac cggcccttag tcatcaagac tggtgtccag

1101  tttaccacga aagtcaggtt gctggtcaaa tttcctgagt tgaattatca

1151  gcttaaaatt aaagtgtgca ttgataaaga ctctggggat gttgctgccc

1201  tcagagggtc tcggaaattt aacattctgg gcacgaacac aaaagtgatg

1251  aacatggagg agtctaacaa cggcagcctg tctgcagagt tcaagcacct

1301  gacccttagg gagcagagat gtgggaatgg aggccgtgcc aattgtgatg

1351  cctccttgat cgtgactgag gagctgcacc tgatcacctt cgagactgag

1401  gtgtaccacc aaggcctcaa gattgaccta gagacccact ccttgccagt

1451  tgtggtgatc tccaacatct gtcagatgcc aaatgcttgg gcatcaatcc

1501  tgtggtataa catgctgacc aataacccca agaacgtgaa cttcttcact

1551  aagccgccaa ttggaacctg ggaccaagtg gccgaggtgc tcagctggca

1601  gttctcgtcc accaccaagc gagggctgag catcgagcag ctgacaacgc

1651  tggctgagaa gctcctaggg cctggtgtga actactcagg gtgtcagatc

1701  acatgggcta aattctgcaa agaaaacatg gctggcaagg gcttctcctt

1751  ctgggtctgg ctagacaata tcatcgacct tgtgaaaaag tatatcttgg

1801  ccctttggaa tgaagggtac atcatgggtt tcatcagcaa ggagcgggag
```

FIG. 15D

19sf6(DNA)

1851 cgggccatcc taagcacaaa gccccgggc accttcctac tgcgcttcag

1901 cgagagcagc aaagaaggag gggtcacttt cacttgggtg gaaaaggaca

1951 tcagtggcaa gacccagatc cagtctgtag agccatacac caagcagcag

2001 ctgaacaaca tgtcatttgc tgaaatcatc atgggctata agatcatgga

2051 tgcgaccaac atcctggtgt ctccacttgt ctacctctac cccgacattc

2101 ccaaggagga ggcatttgga aagtactgta ggcccgagag ccaggagcac

2151 cccgaagccg acccaggtag tgctgccccg tacctgaaga ccaagttcat

2201 ctgtgtgaca ccaacgacct gcagcaatac cattgacctg ccgatgtccc

2251 cccgcacttt agattcattg atgcagtttg gaaataacgg tgaaggtgct

2301 gagccctcag caggagggca gtttgagtcg ctcacgtttg acatggatct

2351 gacctcggag tgtgctacct cccccatgtg aggagctgaa accagaagct

2401 gcagagacgt gacttgagac acctgccccg tgctccaccc ctaagcagcc

2451 gaaccccata tcgtctgaaa ctcctaactt tgtggttcca gatttttttt

2501 tttaatttcc tacttctgct atctttgggc aatctgggca ctttttaaaa

2551 gagagaaatg agtgagtgtg ggtgataaac tgttatgtaa agaggagaga

2601 cctctgagtc tggggatggg gctgagagca gaagggaggc aaaggggaac

2651 acctcctgtc ctgcccgcct gccctccttt ttcagcagct cgggggttgg

2701 ttgttagaca agtgcctcct ggtgcccatg gctacctgtt gccccactct

2751 gtgagctgat accccattct gggaactcct ggctctgcac tttcaacctt

FIG. 15E

19sf6(DNA)

2001 gctaatatcc acatagaagc taggactaag cccaggaggt tcctctttaa

2051 attaaaaaaa aaaaaaaaa

FIG. 16A

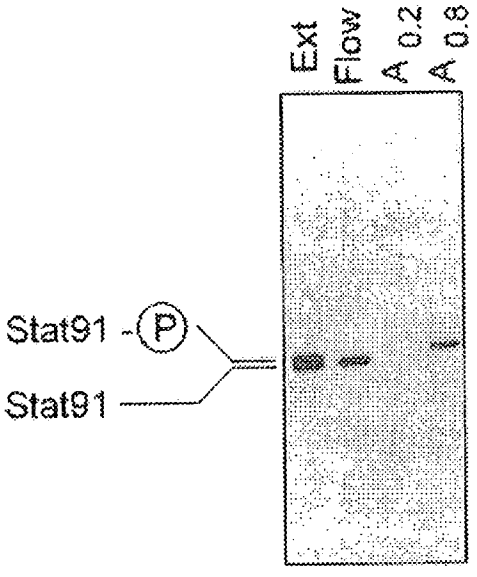


FIG. 16B

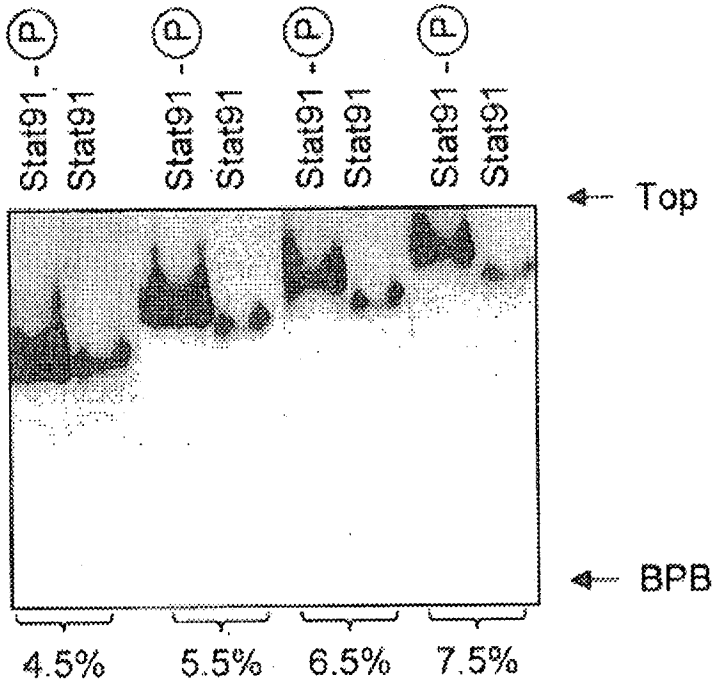


FIG.16C

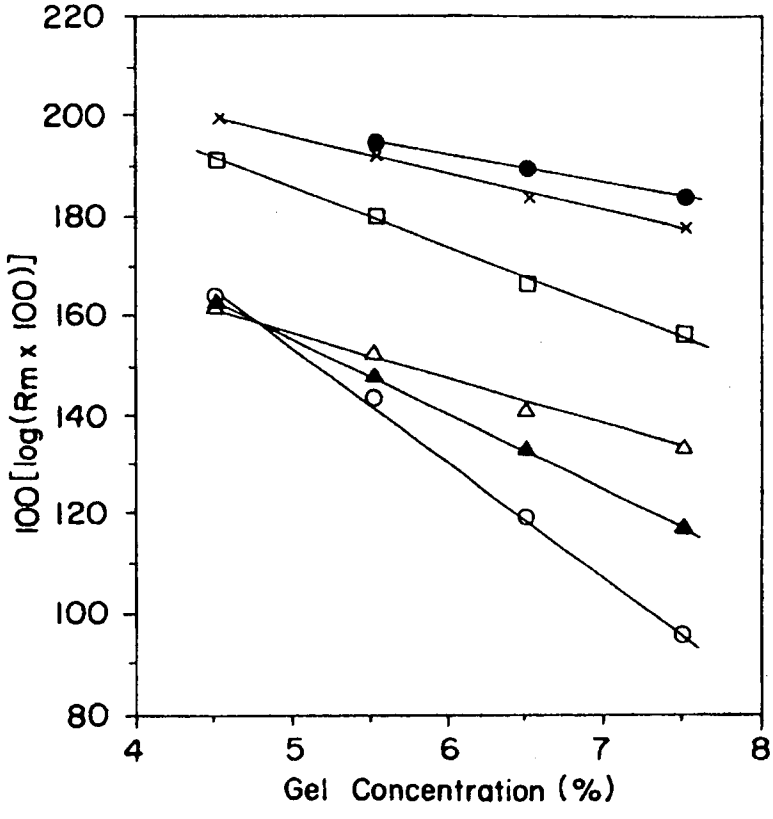


FIG.16D

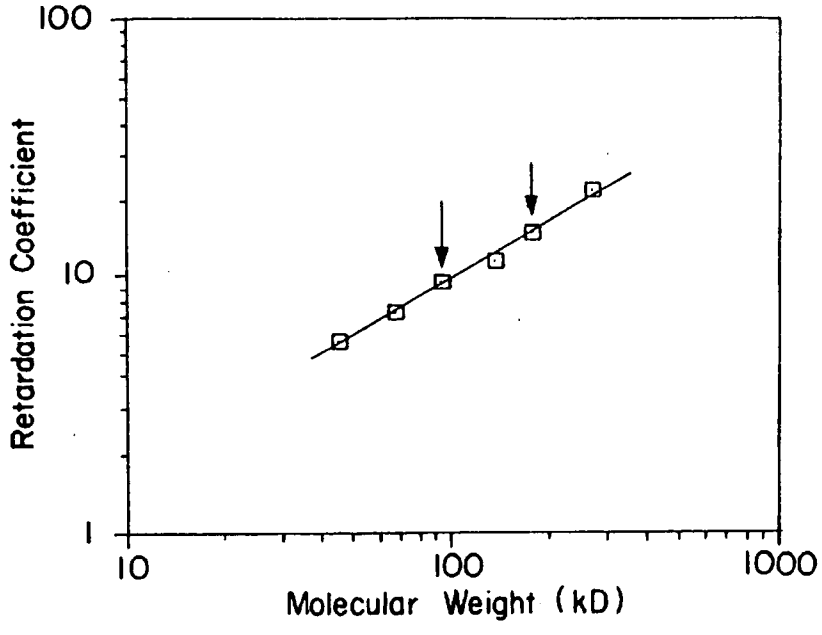


FIG. 17A

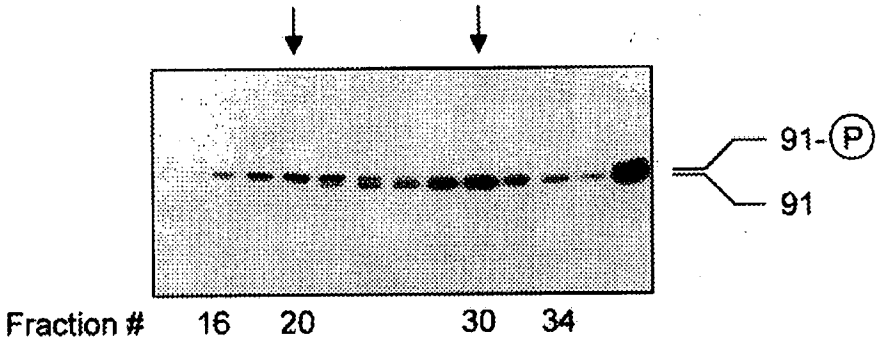


FIG. 17B

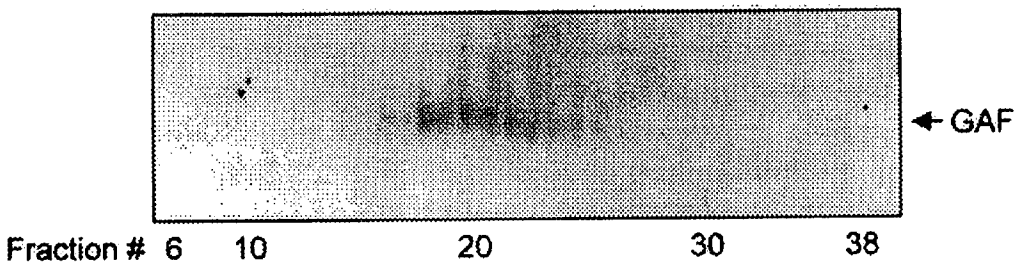


FIG. 17C

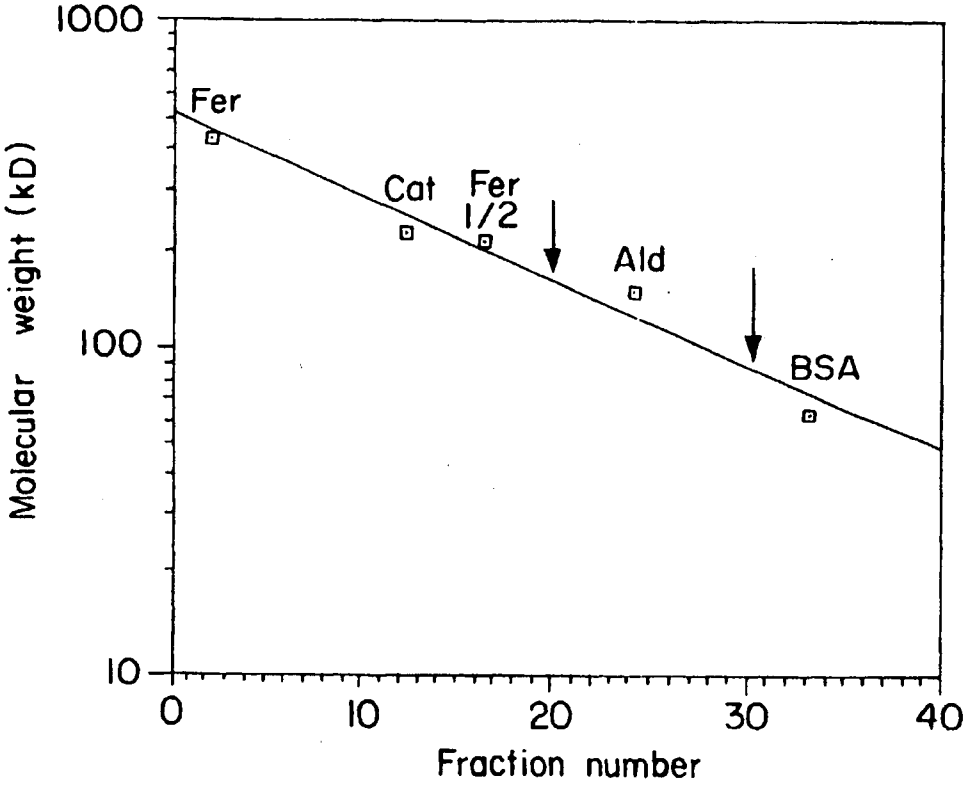

1000
100
10
Molecular weight (kD)
Fer
Cat
Fer 1/2
Ald
BSA
0
10
20
30
40
Fraction number

FIG. 18A

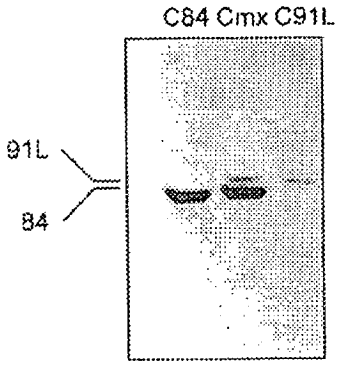


FIG. 18B

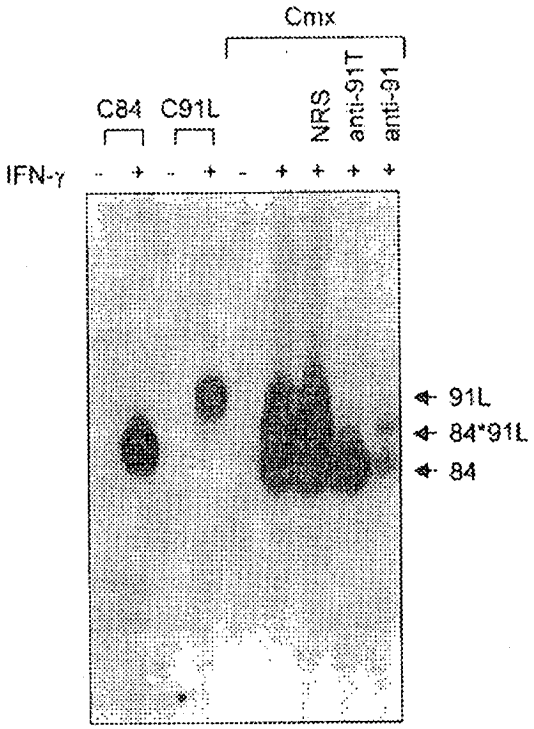


FIG. 19

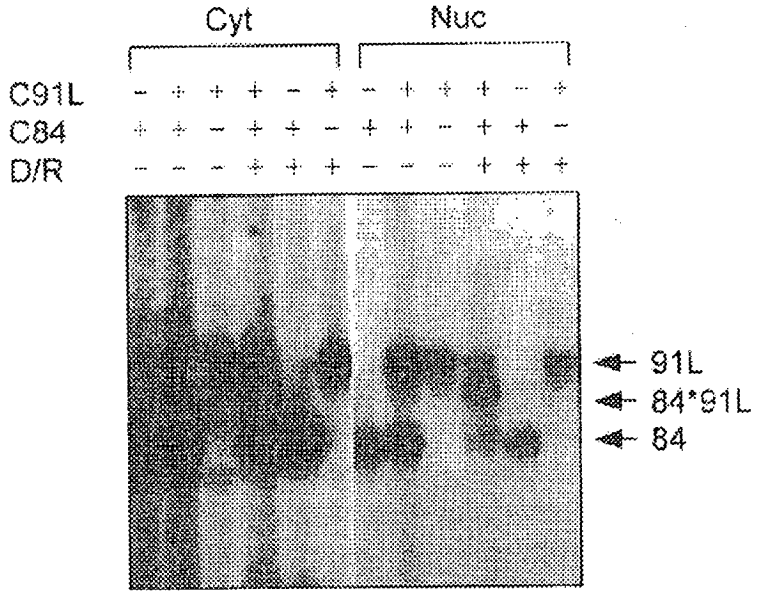

Cyt
Nuc
C91L
C84
D/R
91L
84*91L
84

FIG. 20

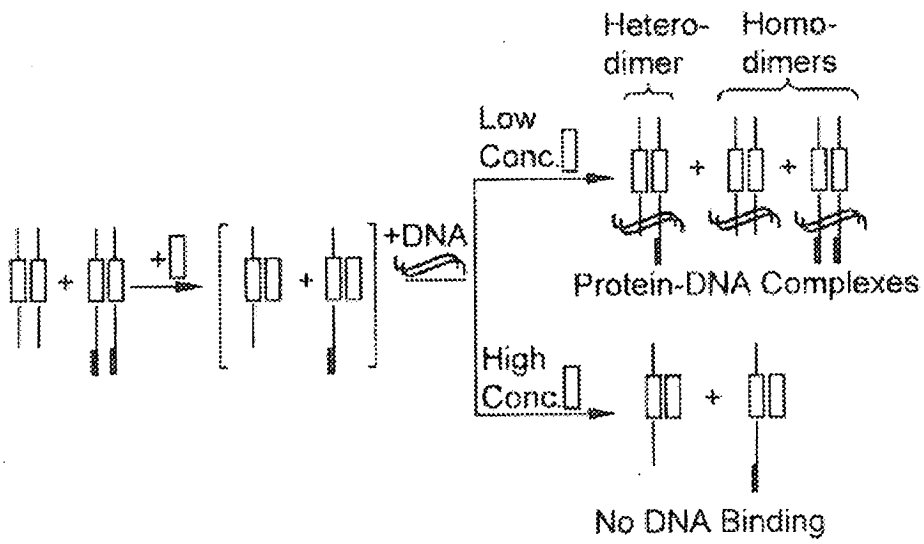

Hetero-dimer
Homo-dimers
Low Conc.
+DNA
Protein-DNA Complexes
High Conc.
No DNA Binding

FIG. 21

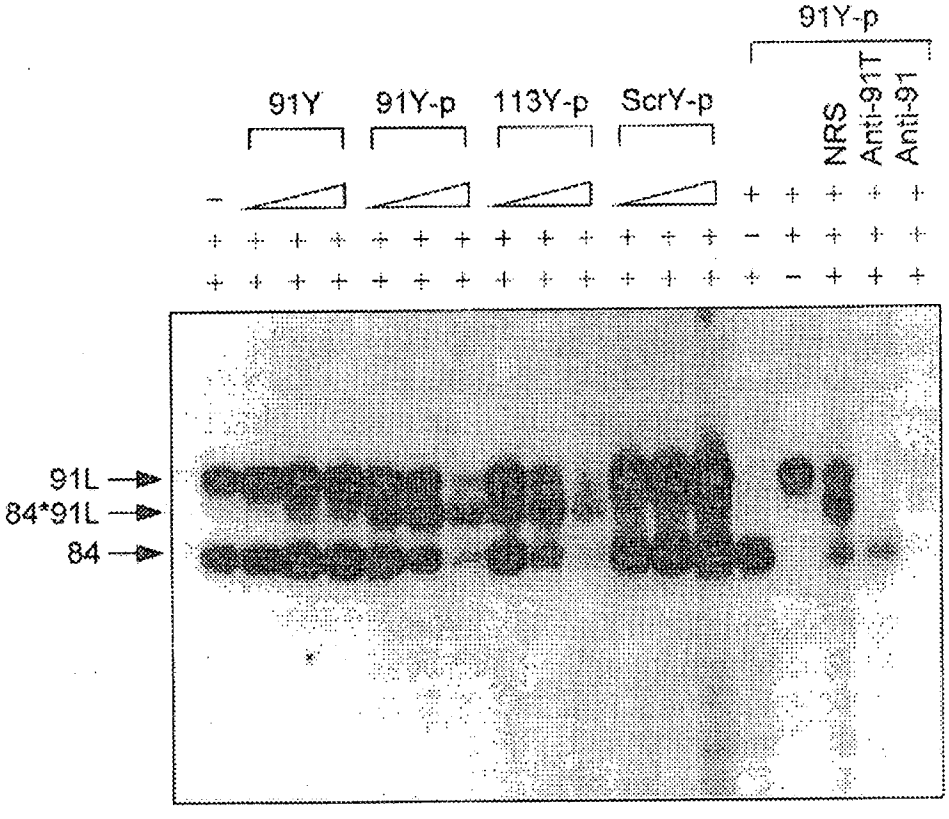


FIG. 22A

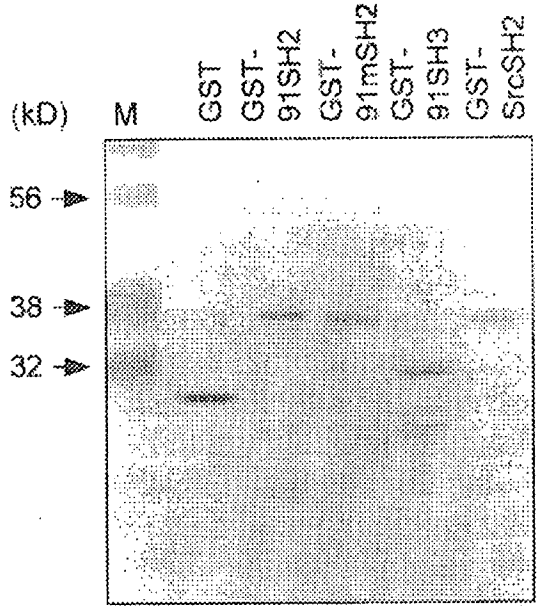


FIG. 22B

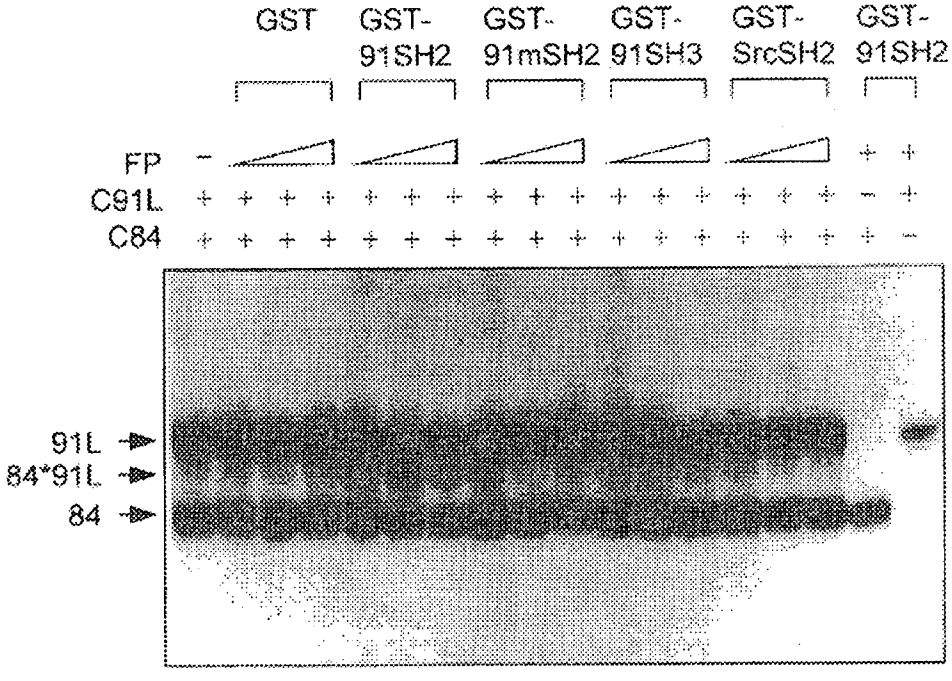


FIG.23A

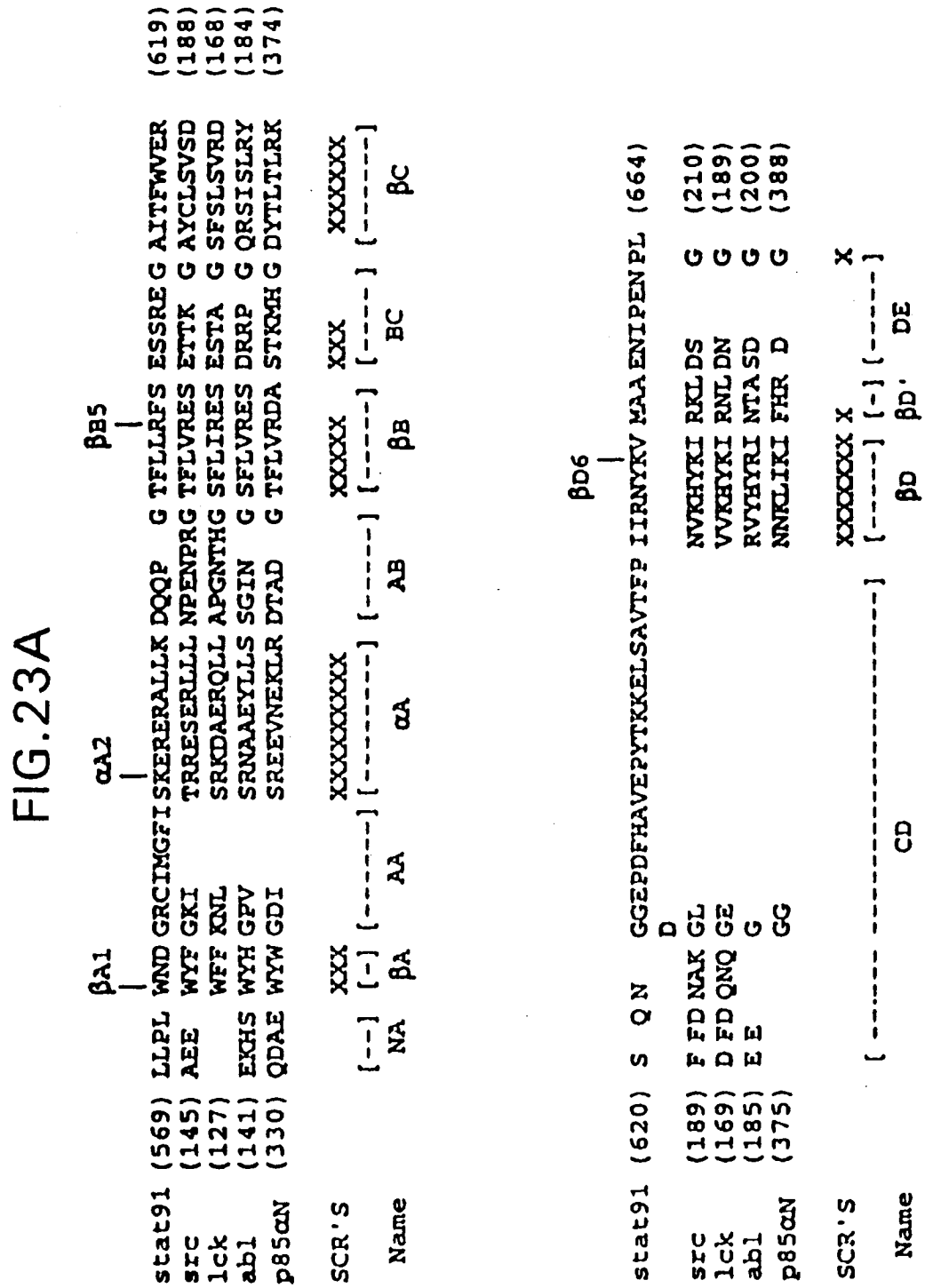
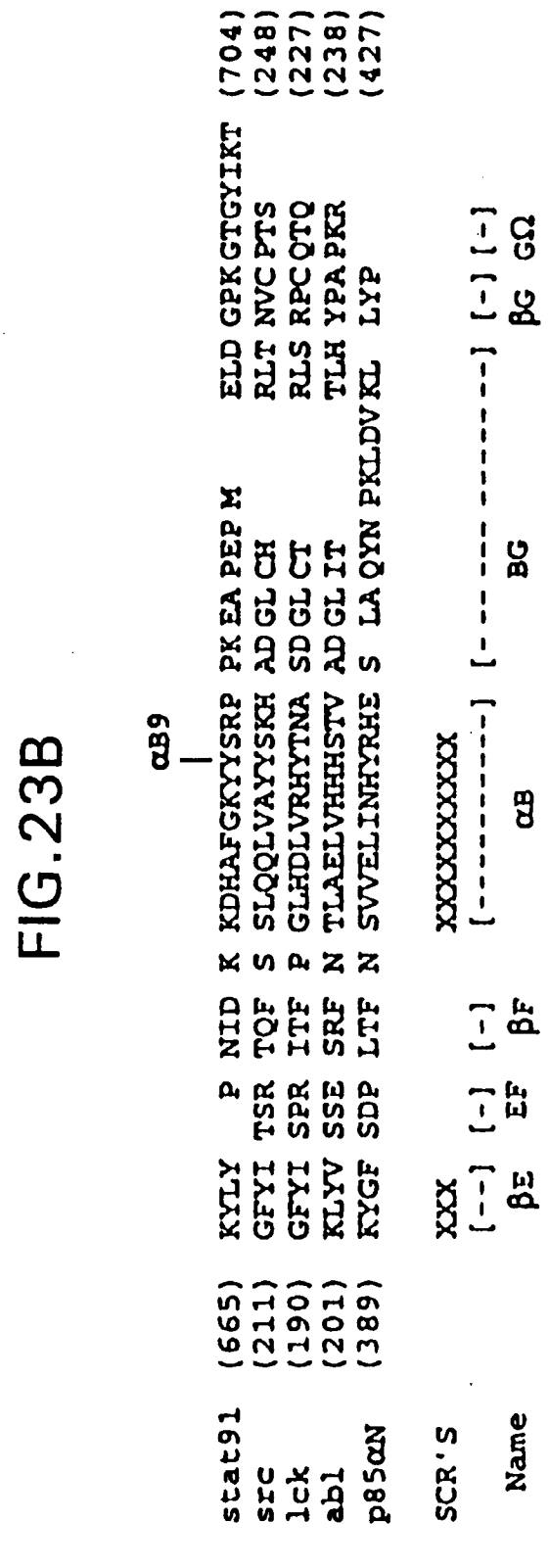

βA1
αA2
βB5
stat91 (569) LLPL WND GRCIMGFI SKERERALLK DQQP G TFLLRFS ESSRE G AITFWVER (619)
src (145) AEE WYF GKI TRRESERLLL NPENPRG TFLVRES ETTK G AYCLSVSD (188)
lck (127) WFF KNL SRKDAERQLL APGNTHG SFLIRES ESTA G SFSLSVRD (168)
abl (141) EKHS WYH GPV SRNAAEYLLS SGIN G SFLVRES DRRP G QRSISLRY (184)
p85αN (330) QDAE WYW GDI SREEVNEKLR DTAD G TFLVRDA STKMH G DYTLTLRK (374)
SCR'S XXX XXXXXXXXX XXXXX XXX XXXXXX
[--] [-] [-------] [---------] [-----] [-----] [----] [------]
Name NA βA AA αA AB βB BC βC
βD6
stat91 (620) S Q N GGEPDFHAVEPYTKKELSAVTFP IIRNYKV MAAENIPENPL (664)
D
src (189) F FD NAK GL NVKHYKI RKLDS G (210)
lck (169) D FD QNQ GE VVKHYKI RNLDN G (189)
abl (185) E E G RVYHYRI NTASD G (200)
p85αN (375) GG NNKLIKI FHRD G (388)
SCR'S XXXXXXX X X
[------ ---------- ----------] [-----] [-] [-----]
Name CD βD βD' DE

FIG.23B


αB9
stat91 (665) KYLY P NID K KDHAFGKYYSRP PK EA PEP M ELD GPK GTGYIKT (704)
src (211) GFYI TSR TQF S SLQQLVAYYSKH AD GL CH RLT NVC PTS (248)
lck (190) GFYI SPR ITF P GLHDLVRHYTNA SD GL CT RLS RPC QTQ (227)
abl (201) KLYV SSE SRF N TLAELVHHHSTV AD GL IT TLH YPA PKR (238)
p85αN (389) KYGF SDP LTF N SVVELINHYRHE S LA QYN PKLDV KL LYP (427)
SCR'S XXX XXXXXXXXXX
(--) (-) (-) (----------) [--- --- --------] [-] [-]
Name βE EF βF αB BG βG GΩ

METHODS OF TESTING DRUGS OR AGENTS THAT MODULATE THE ACTIVITY OF RECEPTOR RECOGNITION FACTORS

CROSS-REFERENCE TO RELATED APPLICATIONS

The present Application is a Continuation-In-Part of U.S. Ser. No. 08/126,588 and U.S. Ser. No. 08/126,595, both filed Sep. 24, 1994, both now abandoned which are both Continuations-In-Part of U.S. Ser. No. 07/980,498, filed Nov. 23, 1992, now abandoned which is a Continuation-In-Part of U.S. Ser. No. 07/854,296, filed Mar. 19, 1992, now abandoned the disclosures of which are hereby incorporated by reference in their entireties. Applicants claim the benefits of these Applications under 35 U.S.C. §120.

RELATED PUBLICATIONS

The Applicants are authors or co-authors of several articles directed to the subject matter of the present invention. (1) Darnell et al., "Interferon-Dependent Transcriptional Activation: Signal Transduction Without Second Messenger Involvement?" *THE NEW BIOLOGIST,* 2(10):1–4, (1990); (2) X. Fu et al., "ISGF3, The Transcriptional Activator Induced by Interferon α, Consists of Multiple Interacting, Polypeptide Chains" *PROC. NATL. ACAD. SCI. USA,* 87:8555–8559 (1990); (3) D. S. Kessler et al., "IFNα Regulates Nuclear Translocation and DNA-Binding Affinity of ISGF3, A Multimeric Transcriptional Activator" *GENES AND DEVELOPMENT,* 4:1753 (1990). All of the above listed articles are incorporated herein by reference.

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to intracellular receptor recognition proteins or factors (i.e. groups of proteins), and to methods and compositions including such factors or the antibodies reactive toward them, or analogs thereof in assays and for diagnosing, preventing and/or treating cellular debilitation, derangement or dysfunction. More particularly, the present invention relates to particular IFN-dependent receptor recognition molecules that have been identified and sequenced, and that demonstrate direct participation in intracellular events, extending from interaction with the liganded receptor at the cell surface to transcription in the nucleus, and to antibodies or to other entities specific thereto that may thereby selectively modulate such activity in mammalian cells.

BACKGROUND OF THE INVENTION

There are several possible pathways of signal transduction that might be followed after a polypeptide ligand binds to its cognate cell surface receptor. Within minutes of such ligand-receptor interaction, genes that were previously quiescent are rapidly transcribed (Murdoch et al., 1982; Larner et al., 1984; Friedman et al., 1984; Greenberg and Ziff, 1984; Greenberg et al., 1985). One of the most physiologically important, yet poorly understood, aspects of these immediate transcriptional responses is their specificity: the set of genes activated, for example, by platelet-derived growth factor (PDGF), does not completely overlap with the one activated by nerve growth factor (NGF) or tumor necrosis factor (TNF) (Cochran et al., 1983; Greenberg et al., 1985; Almendral et al., 1988; Lee et al., 1990). The interferons (IFN) activate sets of other genes entirely. Even IFNα and IFNγ, whose presence results in the slowing of cell growth and in an increased resistance to viruses (Tamm et al., 1987) do not activate exactly the same set of genes (Larner et al., 1984; Friedman et al., 1984; Celis et al., 1987, 1985; Larner et al., 1986).

The current hypotheses related to signal transduction pathways in the cytoplasm do not adequately explain the high degree of specificity observed in polypeptide-dependent transcriptional responses. The most commonly discussed pathways of signal transduction that might ultimately lead to the nucleus depend on properties of cell surface receptors containing tyrosine kinase domains [for example, PDGF, epidermal growth factor (EGF), colony-stimulating factor (CSF), insulin-like growth factor-1 (IGF-1); see Gill, 1990; Hunter, 1990) or of receptors that interact with G-proteins (Gilman, 1987). These two groups of receptors mediate changes in the intracellular concentrations of second messengers that, in turn, activate one of a series of protein phosphokinases, resulting in a cascade of phosphorylations (or dephosphorylations) of cytoplasmic proteins.

It has been widely conjectured that the cascade of phosphorylations secondary to changes in intracellular second messenger levels is responsible for variations in the rates of transcription of particular genes (Bourne, 1988, 1990; Berridge, 1987; Gill, 1990; Hunter, 1990). However, there are at least two reasons to question the suggestion that global changes in second messengers participate in the chain of events leading to specific transcriptional responses dependent on specific receptor occupation by polypeptide ligands.

First, there is a limited number of second messengers (cAMP, diacyl glycerol, phosphoinositides, and Ca^{2+} are the most prominently discussed), whereas the number of known cell surface receptor-ligand pairs of only the tyrosine kinase and G-protein varieties, for example, already greatly outnumbers the list of second messengers, and could easily stretch into the hundreds (Gill, 1990; Hunter, 1990). In addition, since many different receptors can coexist on one cell type at any instant, a cell can be called upon to respond simultaneously to two or more different ligands with an individually specific transcriptional response each involving a different set of target genes. Second, a number of receptors for polypeptide ligands are now known that have neither tyrosine kinase domains nor any structure suggesting interaction with G-proteins. These include the receptors for interleukin-2 (IL-2) (Leonard et al., 1985), IFNα (Uze et al., 1990), IFNγ (Aguet et al., 1988), NGF (Johnson et al., 1986), and growth hormone (Leung et al., 1987). The binding of each of these receptors to its specific ligand has been demonstrated to stimulate transcription of a specific set of genes. For these reasons it seems unlikely that global intracellular fluctuations in a limited set of second messengers are integral to the pathway of specific, polypeptide ligand-dependent, immediate transcriptional responses.

In PCT International Publication No. WO 92/08740 published May 29, 1992 by the applicant herein, the above analysis was presented and it was discovered and proposed that a receptor recognition factor or factors, served in some capacity as a type of direct messenger between liganded receptors at the cell surface and the cell nucleus. One of the characteristics that was ascribed to the receptor recognition factor was its apparent lack of requirement for changes in second messenger concentrations. Continued investigation of the receptor recognition factor through study of the actions of the interferons IFNα and IFNγ has further elucidated the characteristics and structure of the interferon-related factor ISGF-3, and more broadly, the characterization and structure of the receptor recognition factor in a manner that extends beyond earlier discoveries previously described. It is accordingly to the presentation of this

updated characterization of the receptor recognition factor and the materials and methods both diagnostic and therapeutic corresponding thereto that the present disclosure is directed.

SUMMARY OF THE INVENTION

In accordance with the present invention, receptor recognition factors have been further characterized that appear to interact directly with receptors that have been occupied by their ligand on cellular surfaces, and which in turn either become active transcription factors, or activate or directly associate with transcription factors that enter the cells' nucleus and specifically binds on predetermined sites and thereby activates the genes. It should be noted that the receptor recognition proteins thus possess multiple properties, among them: 1) recognizing and being activated during such recognition by receptors; 2) being translocated to the nucleus by an inhibitable process (eg. NaF inhibits translocation); and 3) combining with transcription activating proteins or acting themselves as transcription activation proteins, and that all of these properties are possessed by the proteins described herein.

A further property of the receptor recognition factors (also termed herein signal transducers and activators of transcription—STAT) is dimerization to form homodimers or heterodimers upon activation by phosphorylation of tyrosine. In a specific embodiment, infra, Stat91 and Stat84 form homodimers and a Stat91–Stat84 heterodimer. Accordingly, the present invention is directed to such dimers, which can form spontaneously by phophorylation of the STAT protein, or which can be prepared synthetically by chemically cross-linking two like or unlike STAT proteins.

The receptor recognition factor is proteinaceous in composition and is believed to be present in the cytoplasm. The recognition factor is not demonstrably affected by concentrations of second messengers, however does exhibit direct interaction with tyrosine kinase domains, although it exhibits no apparent interaction with G-proteins. More particularly, as is shown in a co-pending, co-owned application entitled "INTERFERON-ASSOCIATED RECEPTOR RECOGNITION FACTORS, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF USE THEREOF," filed on even date herewith, the 91 kD human interferon (IFN)-γ factor, represented by SEQ ID NO:4 directly interacts with DNA after acquiring phosphate on tyrosine located at position 701 of the amino acid sequence.

The recognition factor is now known to comprise several proteinaceous substituents, in the instance of IFNα and IFNγ. Particularly, three proteins derived from the factor ISGF-3 have been successfully sequenced and their sequences are set forth in FIG. **1** (SEQ ID NOS:1, 2), FIG. **2** (SEQ ID NOS:3, 4) and FIG. **3** (SEQ. ID NOS.5, 6) herein. Additionally, a murine gene encoding the 91 kD protein (SEQ ID NO:4) has been identified and sequenced. The nucleotide sequence (SEQ ID NO:7) and deduced amino acid sequence (SEQ ID NO:8) are shown in FIGS. **13A–13C**.

In a further embodiment, murine genes encoding homologs of the recognition factor have been succefully sequenced and cloned into plasmids. A gene in plasmid 13sf1 has the nucleotide sequence (SEQ ID NO:9) and deduced amino acid sequence (SEQ ID NO:10) as shown in FIGS. **14A–14C**. A gene in plasmid 19sf6 has the nucleotide sequence (SEQ ID NO:11) and deduced amino acid sequence (SEQ ID NO:12) shown in FIGS. **15A–15C**.

It is particularly noteworthy that the protein sequence of FIG. **1** (SEQ ID NO:2) and the sequence of the proteins of FIGS. **2** (SEQ ID NO:4) and **3** (SEQ ID NO:6) derive, respectively, from two different but related genes. Moreover, the protein sequence of FIG. **13** (SEQ ID NO:8) derives from a murine gene that is analogous to the gene encoding the protein of FIG. **2** (SEQ ID NO:4). Of further note is that the protein sequences of FIGS. **14** (SEQ ID NO:10) and **15** (SEQ ID NO:12) derive from two genes that are different from, but related to, the protein of FIG. **13** (FIG ID NO:8). It is clear from these discoveries that a family of genes exists, and that further family members likewise exist. Accordingly, as demonstrated herein, by use of hybridization techniques, additional such family members will be found.

Further, the capacity of such family members to function in the manner of the receptor recognition factors disclosed, herein may be assessed by determining those ligand that cause the phosphorylation of the particular family members.

In its broadest aspect, the present invention extends to a receptor recognition factor implicated in the transcriptional stimulation of genes in target cells in response to the binding of a specific polypeptide ligand to its cellular receptor on said target cell, said receptor recognition factor having the following characteristics:

a) apparent direct interaction with the ligand-bound receptor complex and activation of one or more transcription factors capable of binding with a specific gene;

b) an activity demonstrably unaffected by the presence or concentration of second messengers;

c) direct interaction with tyrosine kinase domains; and

d) a perceived absence of interaction with G-proteins.

In a further aspect, the receptor recognition (STAT) protein forms a dimer upon activation by phosphorylation.

In a specific example, the receptor recognition factor represented by SEQ ID NO:4 possesses the added capability of acting as a transcription factor and, in particular, as a DNA binding protein in response to interferon-γ stimulation. This discovery presages an expanded role for the proteins in question, and other proteins and like factors that have heretofore been characterized as receptor recognition factors. It is therefore apparent that a single factor may indeed provide the nexus between the liganded receptor at the cell surface and direct participation in DNA transcriptional activity in the nucleus. This pleiotypic factor has the following characteristics:

a) It interacts with an interferon-γ-bound receptor kinase complex;

b) It is a tyrosine kinase substrate; and

c) When phosphorylated, it serves as a DNA binding protein.

More particularly, the factor represented by SEQ ID NO:4 is interferon-dependent in its activity and is responsive to interferon stimulation, particularly that of interferon-γ. It has further been discovered that activation of the factor represented by SEQ: ID NO:4 requires phosphorylation of tyrosine-701 of the protein, and further still that tyrosine phosphorylation requires the presence of a functionally active SH2 domain in the protein. Preferably, such SH2 domain contains an amino acid residue corresponding to an arginine at position 602 of the protein.

In a still further aspect, the present invention extends to a receptor recognition factor interactive with a liganded interferon receptor, which receptor recognition factor possesses the following characteristics:

a) it is present in cytoplasm;

b) it undergoes tyrosine phosphorylation upon treatment of cells with IFNα or IFNγ;

c) it activates transcription of an interferon stimulated gene;

d) it stimulates either an ISRE-dependent or a gamma activated site (GAS)-dependent transcription in vivo;

e) it interacts with IFN cellular receptors, and

f) it undergoes nuclear translocation upon stimulation of the IFN cellular receptors with IFN.

The factor of the invention represented by SEQ ID NO:4 appears to act in similar fashion to an earlier determined site-specific DNA binding protein that is interferon-γ dependent and that has been earlier called the γ activating factor (GAF). Specifically, interferon-γ-dependent activation of this factor occurs without new protein synthesis and appears within minutes of interferon-γ treatment, achieves maximum extent between 15 and 30 minutes thereafter, and then disappears after 2–3 hours. These further characteristics of identification and action assist in the evaluation of the present factor for applications having both diagnostic and therapeutic significance.

In a particular embodiment, the present invention relates to all members of the herein disclosed family of receptor recognition factors except the 91 kD protein factors, specifically the proteins whose sequences are represented by one or more of SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8.

The present invention also relates to a recombinant DNA molecule or cloned gene, or a degenerate variant thereof, which encodes a receptor recognition factor, or a fragment thereof, that possesses a molecular weight of about 113 kD and an amino acid sequence set forth in FIG. **1** (SEQ ID NO:2); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 113 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA sequence shown in FIG. **1** (SEQ ID NO:1). In another embodiment, the receptor recognition factor has a molecular weight of about 91 kD and the amino acid sequence set forth in FIG. **2** (SEQ ID NO:4) or FIG. **13** (SEQ ID NO:8); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 91 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqnece shown in FIG. **2** (SEQ ID NO:3) or FIG. **13** (SEQ ID NO:8). In yet a further embodiment, the receptor recognition factor has a molecular weight of about 84 kD and the amino acid sequence set forth in FIG. **3** (SEQ ID NO:6); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 84 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqnece shown in FIG. **3** (SEQ ID NO:5). In yet another embodiment, the receptor recognition factor has an amino acid sequence set forth in FIG. **14** (SEQ ID NO:10); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding such receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqnece shown in FIG. **14** (SEQ ID NO:9). In still another embodiment, the receptor recognition factor has an amino acid sequence set forth in FIG. **15** (SEQ ID NO:12); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding such receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqnece shown in FIG. **15** (SEQ ID NO:11).

The human and murine DNA sequences of the receptor recognition factors of the present invention or portions thereof, may be prepared as probes to screen for complementary sequences and genomic clones in the same or alternate species. The present invention extends to probes so prepared that may be provided for screening cDNA and genomic libraries for the receptor recognition factors. For example, the probes may be prepared with a variety of known vectors, such as the phage λ vector. The present invention also includes the preparation of plasmids including such vectors, and the use of the DNA sequences to construct vectors expressing antisense RNA or ribozymes which would attack the mRNAs of any or all of the DNA sequences set forth in FIGS. **1**, **2**, **3**, **13**, **14** and **15** (SEQ ID NOS:1, 3, 5, 7, 9, and 11, respectively). Correspondingly, the preparation of antisense RNA and ribozymes are included herein.

The present invention also includes receptor recognition factor proteins having the activities noted herein, and that display the amino acid sequences set forth and described above and selected from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10 and SEQ ID NO:12.

In a further embodiment of the invention, the full DNA sequence of the recombinant DNA molecule or cloned gene so determined may be operatively linked to an expression control sequence which may be introduced into an appropriate host. The invention accordingly extends to unicellular hosts transformed with the cloned gene or recombinant DNA molecule comprising a DNA sequence encoding the present receptor recognition factor(s), and more particularly, the complete DNA sequence determined from the sequences set forth above and in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9 and SEQ ID NO:11.

According to other preferred features of certain preferred embodiments of the present invention, a recombinant expression system is provided to produce biologically active animal or human receptor recognition factor.

The concept of the receptor recognition factor contemplates that specific factors exist for correspondingly specific ligands, such as tumor necrosis factor, nerve growth factor and the like, as described earlier. Accordingly, the exact structure of each receptor recognition factor will understandably vary so as to achieve this ligand and activity specificity. It is this specificity and the direct involvement of the receptor recognition factor in the chain of events leading to gene activation, that offers the promise of a broad spectrum of diagnostic and therapeutic utilities.

The present invention naturally contemplates several means for preparation of the recognition factor, including as illustrated herein known recombinant techniques, and the invention is accordingly intended to cover such synthetic preparations within its scope. The isolation of the cDNA amino acid sequences disclosed herein facilitates the reproduction of the recognition factor by such recombinant techniques, and accordingly, the invention extends to expression vectors prepared from the disclosed DNA sequences for expression in host systems by recombinant DNA techniques, and to the resulting transformed hosts.

The invention includes an assay system for screening of potential drugs effective to modulate transcriptional activity of target mammalian cells by interrupting or potentiating the recognition factor or factors. In one instance, the test drug could be administered to a cellular sample with the ligand that activates the receptor recognition factor, or an extract containing the activated recognition factor, to determine its effect upon the binding activity of the recognition factor to any chemical sample (including DNA), or to the test drug, by comparison with a control.

The assay system could more importantly be adapted to identify drugs or other entities that are capable of binding to the receptor recognition and/or transcription factors or

proteins, either in the cytoplasm or in the nucleus, thereby inhibiting or potentiating transcriptional activity. Such assay would be useful in the development of drugs that would be specific against particular cellular activity, or that would potentiate such activity, in time or in level of activity. For example, such drugs might be used to modulate cellular response to shock, or to treat other pathologies, as for example, in making IFN more potent against cancer.

In yet a further embodiment, the invention contemplates antagonists of the activity of a receptor recognition factor (STAT). In particular, an agent or molecule that inhibits dimerization (homodimerization or heterodimerization) can be used to block transcription activation effected by an acitvated, phosphorylated STAT protein. In a specific embodiment, the antagonist can be a peptide having the sequence of a portion of an SH2 domain of a STAT protein, or the phophotyrosine domaine of a STAT protein, or both. If the peptide contains both regions, preferably the regions are located in tandem, more preferably with the SH2 domain portion N-terminal to the phosphotyrosine portion. In a specific example, infra, such peptides are shown to be capable of disrupting dimerization of STAT proteins.

One of the characteristics of the present receptor recognition factors is their participation in rapid phosphorylation and dephosphorylation during the course of and as part of their activity. Significantly, such phosphorylation takes place in an interferon-dependent manner and within a few minutes in the case of the ISGF-3 proteins identified herein, on the tyrosine residues defined thereon. This is strong evidence that the receptor recognition factors disclosed herein are the first true substrates whose intracellular function is well understood and whose intracellular activity depends on tyrosine kinase phosphorylation. In particular, the addition of phosphate to the tyrosine of a transcription factor is novel. This suggests further that tyrosine kinase takes direct action in the transmission of intracellular signals to the nucleus, and does not merely serve as a promoter or mediator of serine and/or serinine kinase activity, as has been theorized to date. Also, the role of the factor represented by SEQ ID NO:2 in its activated phosphorylated form suggests possible independent therapeutic use for this activated form. Likewise, the role of the factor as a tyrosine kinase substrate suggests its interaction with kinase in other theatres apart from the complex observed herein.

The diagnostic utility of the present invention extends to the use of the present receptor recognition factors in assays to screen for tyrosine kinase inhibitors.

Because the activity of the receptor recognition-transcriptional activation proteins described herein must maintain tyrosine phosphorylation, they can and presumably are dephosphorylated by specific tyrosine phosphatases. Blocking of the specific phosphatase is therefore an avenue of pharmacological intervention that would potentiate the activity of the receptor recognition proteins.

The present invention likewise extends to the development of antibodies against the receptor recognition factor(s), including naturally raised and recombinantly prepared antibodies. For example, the antibodies could be used to screen expression libraries to obtain the gene or genes that encode the receptor recognition factor(s). Such antibodies could include both polyclonal and monoclonal antibodies prepared by known genetic techniques, as well as bi-specific (chimeric) antibodies, and antibodies including other functionalities suiting them for additional diagnostic use conjunctive with their capability of modulating transcriptional activity.

In particular, antibodies against specifically phosphorylated factors can be selected and are included within the scope of the present invention for their particular ability in following activated protein. Thus, activity of the recognition factors or of the specific polypeptides believed to be causally connected thereto may therefore be followed directly by the assay techniques discussed later on, through the use of an appropriately labeled quantity of the recognition factor or antibodies or analogs thereof.

Thus, the receptor recognition factors, their analogs and/or analogs, and any antagonists or antibodies that may be raised thereto, are capable of use in connection with various diagnostic techniques, including immunoassays, such as a radioimmunoassay, using for example, an antibody to the receptor recognition factor that has been labeled by either radioactive addition, reduction with sodium borohydride, or radioiodination.

In an immunoassay, a control quantity of the antagonists or antibodies thereto, or the like may be prepared and labeled with an enzyme, a specific binding partner and/or a radioactive element, and may then be introduced into a cellular sample. After the labeled material or its binding partner(s) has had an opportunity to react with sites within the sample, the resulting mass may be examined by known techniques, which may vary with the nature of the label attached. For example, antibodies against specifically phosphorylated factors may be selected and appropriately employed in the exemplary assay protocol, for the purpose of following activated protein as described above.

In the instance where a radioactive label, such as the isotopes ^{3}H, ^{14}C, ^{32}P, ^{35}S, ^{36}Cl, ^{51}Cr, ^{57}Co, ^{58}Co, ^{59}Fe, ^{90}Y, ^{125}I, ^{131}I, and ^{186}Re are used, known currently available counting procedures may be utilized. In the instance where the label is an enzyme, detection may be accomplished by any of the presently utilized colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques known in the art.

The present invention includes an assay system which may be prepared in the form of a test kit for the quantitative analysis of the extent of the presence of the recognition factors, or to identify drugs or other agents that may mimic or block their activity. The system or test kit may comprise a labeled component prepared by one of the radioactive and/or enzymatic techniques discussed herein, coupling a label to the recognition factors, their agonists and/or antagonists, and one or more additional immunochemical reagents, at least one of which is a free or immobilized ligand, capable either of binding with the labeled component, its binding partner, one of the components to be determined or their binding partner(s).

In a further embodiment, the present invention relates to certain therapeutic methods which would be based upon the activity of the recognition factor(s), its (or their) subunits, or active fragments thereof, or upon agents or other drugs determined to possess the same activity. A first therapeutic method is associated with the prevention of the manifestations of conditions causally related to or following from the binding activity of the recognition factor or its subunits, and comprises administering an agent capable of modulating the production and/or activity of the recognition factor or subunits thereof, either individually or in mixture with each other in an amount effective to prevent the development of those conditions in the host. For example, drugs or other binding partners to the receptor recognition/transcription factors or proteins may be administered to inhibit or potentiate transcriptional activity, as in the potentiation of interferon in cancer therapy. Also, the blockade of the action of specific tyrosine phosphatases in the dephosphorylation of activated (phosphorylated) recognition/transcription factors

or proteins presents a method for potentiating the activity of the receptor recognition factor or protein that would concomitantly potentiate therapies based on receptor recognition factor/protein activation.

More specifically, the therapeutic method generally referred to herein could include the method for the treatment of various pathologies or other cellular dysfunctions and derangements by the administration of pharmaceutical compositions that may comprise effective inhibitors or enhancers of activation of the recognition factor or its subunits, or other equally effective drugs developed for instance by a drug screening assay prepared and used in accordance with a further aspect of the present invention. For example, drugs or other binding partners to the receptor recognition/transcription factor or proteins, as represented by SEQ ID NO:2, may be administered to inhibit or potentiate transcriptional activity, as in the potentiation of interferon in cancer therapy. Also, the blockade of the action of specific tyrosine phosphatases in the dephosphorylation of activated (phosphorylated) recognition/transcription factor or protein presents a method for potentiating the activity of the receptor recognition factor or protein that would concomitantly potentiate therapies based on receptor recognition factor/protein activation. Correspondingly, the inhibition or blockade of the activation or binding of the recognition/transcription factor would affect MHC Class II expression and consequently, would promote immunosuppression. Materials exhibiting this activity, as illustrated later on herein by staurosporine, may be useful in instances such as the treatment of autoimmune diseases and graft rejection, where a degree of immunosuppression is desirable.

In particular, the proteins of ISGF-3 whose sequences are presented in SEQ ID NOS:2, 4, 6, 8, 10 or 12 herein, their antibodies, agonists, antagonists, or active fragments thereof, could be prepared in pharmaceutical formulations for administration in instances wherein interferon therapy is appropriate, such as to treat chronic viral hepatitis, hairy cell leukemia, and for use of interferon in adjuvant therapy. The specificity of the receptor proteins hereof would make it possible to better manage the aftereffects of current interferon therapy, and would thereby make it possible to apply interferon as a general antiviral agent.

Accordingly, it is a principal object of the present invention to provide a receptor recognition factor and its subunits in purified form that exhibits certain characteristics and activities associated with transcriptional promotion of cellular activity.

It is a further object of the present invention to provide antibodies to the receptor recognition factor and its subunits, and methods for their preparation, including recombinant means.

It is a further object of the present invention to provide a method for detecting the presence of the receptor recognition factor and its subunits in mammals in which invasive, spontaneous, or idiopathic pathological states are suspected to be present.

It is a further object of the present invention to provide a method and associated assay system for screening substances such as drugs, agents and the like, potentially effective in either mimicking the activity or combating the adverse effects of the recognition factor and/or its subunits in mammals.

It is a still further object of the present invention to provide a method for the treatment of mammals to control the amount or activity of the recognition factor or subunits thereof, so as to alter the adverse consequences of such presence or activity, or where beneficial, to enhance such activity.

It is a still further object of the present invention to provide a method for the treatment of mammals to control the amount or activity of the recognition factor or its subunits, so as to treat or avert the adverse consequences of invasive, spontaneous or idiopathic pathological states.

It is a still further object of the present invention to provide pharmaceutical compositions for use in therapeutic methods which comprise or are based upon the recognition factor, its subunits, their binding partner(s), or upon agents or drugs that control the production, or that mimic or antagonize the activities of the recognition factors.

Other objects and advantages will become apparent to those skilled in the art from a review of the ensuing description which proceeds with reference to the following illustrative drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** depicts the full receptor recognition factor nucleic acid sequence and the deduced amino acid sequence derived for the ISGF-3α gene defining the 113 kD) protein. The nucleotides are numbered from 1 to 2553 (SEQ ID NO:1), and the amino acids are numbered from 1 to 851 (SEQ ID NO:2).

FIG. **2** depicts the full receptor recognition factor nucleic acid sequence and the deduced amino acid sequence derived for the ISGF-3α gene defining the 91 kD protein. The nucleotides are numbered from 1 to 3943 (SEQ ID NO:3), and the amino acids are numbered from 1 to 750 (SEQ ID NO:4).

FIG. **3** depicts the full receptor recognition factor nucleic acid sequence and the deduced amino acid sequence derived for the ISGF-3α gene defining the 84 kD protein. The nucleotides are numbered from 1 to 2166 (SEQ ID NO:5), and the amino acids are numbered from 1 to 712 (SEQ ID NO:6).

FIG. **4** shows the purification of ISGF-3. The left-hand portion of the Figure shows the purification of ISGF-3 demonstrating the polypeptides present after the first oligonucleotide affinity column (lane 3) and two different preparations after the final chromatography step (Lanes 1 and 2). The left most lane contains protein size markers (High molecular weight, Sigma). ISGF-3 component proteins are indicated as 113 kD, 91 kD, 84 kD, and 48 kD [Kessler et al., GENES & DEV., 4 (1990); Levy et al., THE EMBO. J., 9 (1990)]. The right-hand portion of the Figure shows purified ISGF-3 from $2–3\times10^{11}$ cells was electroblotted to nitrocellulose after preparations 1 and 2 (Lanes 1 and 2) had been pooled and separated on a 7.5% SDS polyacrylamide gel. ISGF-3 component proteins are indicated. The two lanes on the right represent protein markers (High molecular weight, and prestained markers, Sigma).

FIGS. ***5a–5b*** generally presents the results of Northern Blot analysis for the 91/84 kD peptides. FIG. ***5a*** presents restriction maps for cDNA clones E4 (top map) and E3 (bottom map) showing DNA fragments that were radiolabeled as probes (probes A–D). FIG. ***5b*** comprises Northern blots of cytoplasmic HeLa RNA hybridized with the indicated probes. The 4.4 and 3.1 KB species as well as the 28S and 18S rRNA bands are indicated.

FIG. **6** depicts the conjoint protein sequence of the 91 kD (SEQ ID NO:4) and 84 kD (SEQ ID NO:6) proteins of ISGF-3. One letter amino acid code is shown for the open reading frame from clone E4, (encoding the 91 kD protein). The 84 kD protein, encoded by a different cDNA (E3), has the identical sequence but terminates after amino acid 712, as indicated. Tryptic peptides t19, t13a, and t13b from the 91

kD protein are indicated. The sole recovered tryptic peptide from the 84 kD protein, peptide t27, was wholly contained within peptide t19 as indicated.

FIGS. **7***a–***7**presents the results of Western blot and antibody shift analyses.

a) Highly purified ISGF-3, fractionated on a 7.0% SDS polyacrylamide gel, was probed with antibodies a42 (amino acids 597–703); a55 (amino acids 2–59); and a57 (amino acids 705–739) in a Western blot analysis. The silver stained part of the gel (lanes a, b, and c) illustrates the location of the ISGF-3 component proteins and the purity of the material used in Western blot: Lane a) Silver stain of protein sample used in all the Western blot experiments (immune and preimmune). Lane b) Material of equal purity to that shown in FIG. **4**, for clearer identification of the ISGF-3 proteins. Lane c) Size protein markers indicated.

b) Antibody interference of the ISGF-3 shift complex; Lane a) The complete ISGF-3 and the free ISGF-3γ component shift with partially purified ISGF-3 are marked; Lane b) Competition with a 100 fold excess of cold ISRE oligonucleotide. Lane c) Shift complex after the addition of 1 ml of preimmune serum to a 12.5 μl shift reaction. Lanes d and e)—Shift complex after the addition of 1 μl of a 1:10 dilution or 1 ml of undiluted a42 antiserum to a 12.5 μl shift reaction.

Methods

Antibodies a42, a55 and a57 were prepared by injecting approximately 500 mgm of a fusion protein prepared in *E. coli* using the GE3-3X vector [Smith et al., *GENE,* 67 (1988)]. Rabbits were bled after the second boost and serum prepared.

For Western blots highly purified ISGF-3 was separated on a 7% SDS polyacrylamide gel and electroblotted to nitrocellulose. The filter was incubated in blocking buffer ("blotto"), cut into strips and probed with specific antiserum and preimmune antiserum diluted 1:500. The immune complexes were visualized with the aid of an ECL kit (Amersham). Shift analyses were performed as previously described [Levy et al., *GENES & DEV.,* 2 (1988); Levy et al., *GENES & DEV.,* 3 (1989)] in a 4.5% polyacrylamide gel.

FIG. **8** presents the full length amino acid sequence of 113 kD protein components of ISGF-3α (SEQ ID NO:2) and alignment of conserved amino acid sequences between the 113 kD and 91/84 kD proteins (SEQ ID NOS:4 AND 6).

A. Polypeptide sequences (A–E) derived from protein micro-sequencing of purified 113 kD protein (see accompanying paper) are underlined. Based on peptide E, we designed a degenerate oligonucleotide, AAT/CACIGAA/GCCIATGGAA/GATT/CATT (SEQ ID NO:13), which was used to screen a cDNA library [Pine et la., *MOL. CELL. BIOL.,* 10 (1990)] basically as described [Norman et al., *CELL,* 55 (1988)]. Briefly, the degenerate oligonucleotides were labeled by 32P-γ-ATP by polynucleotide kinase, hybridizations were carried out overnight at 40° C. in 6×SSTE (0.9 M NaCl, 60 mM Tris-HCl [pH 7.9] 6 mM EDTA), 0.1% SDS, 2 mM $Na_2P_5O_7$, 6 mM KH_2PO_4 in the presence of 100 mg/ml salmon sperm DNA sperm and 10×Denhardt's solution [Maniatis et al., *MOLECULAR CLONING, A LABORATORY MANUAL* (Cold Spring Harbor Lab., 1982)]. The nitrocellulose filters then were washed 4×10 min. with the same hybridization conditions without labeled probe and salmon sperm DNA. Autoradiography was carried out at −80° C. with intensifying screen for 48 hrs. A PCR product was obtained later by the same method described for the 91/84 kD sequences, by using oligonucleotides designed according polypeptide D and E. The sequence of this PCR product was identical to a region in clone f11. The full length of 113 kD protein contains 851 amino acids. Three major helices in the N-terminal region were predicted by the methods of both Chou and Fasman [Chou et al., *ANN. REV. BIOCHEM.,* 47 (1978)] and Gamier et al [Gamier et al., *J. MOL. BIOL.,* 12 (1978)] and are shown in shadowed boxes. At the C-terminal end, a highly negative charged domain was found. All negative charged residues are blackened and positive charged residues shadowed. The five polypeptides that derived from protein microscreening [Aebersold et al., *PROC. NATL. ACAD. SCI. USA,* 87 (1987)] are underlined.

B) Comparison of amino acid sequences of 113 kD and 91/84 kD protein shows a 42% identical amino acid residues in the overlapping 715 amino acid sequence shown. In the middle helix region four leucine and one valine heptad repeats were identified in both 113 and 91/84 kD protein (the last leucine in 91/84 kD is not exactly preserved as heptad repeats). When a heligram structure was drawn this helix is amphipathic (not shown). Another notable feature of this comparison is several tyrosine residues that are conserved in both proteins near their ends.

FIG. **9** shows the in vitro transcription and translation of 113 kD and 91 kD cDNA and a Northern blot analysis with 113 kD cDNA probe.

a) The full length cDNA clones of 113 and 91 kD protein were transcribed in vitro and transcribed RNAs was translated in vitro with rabbit lenticulate lysate (Promega; conditions as described in the Promega protocol). The mRNA of BMV (Promega) was simultaneously translated as a protein size marker. The 113 cDNA yielded a translated product about 105 kD and the 91 cDNA yielded a 86 kD product.

b) When total cytoplasmic mRNAs isolated from super-induced HeLa cells were utilized, a single 4.8 KB mRNA band was observed with a cDNA probe coding for C-end of 113 kD protein in a Northern blot analysis [Nielsch et al., *The EMBO. J.,* 10 (1991)].

FIG. **10**(A) presents the results of Western blot analysis confirming the identity of the 113 kD protein. An antiserum raised against a polypeptide segment [Harlow et al., *ANTIBODIES; A LABORATORY MANUAL* (Cold Spring Harbor Lab., 1988)] from amino acid 500 to 650 of 113 kD protein recognized specifically a 113 kD protein in a protein Western blot analysis. The antiserum recognized a band both in a highly purified ISGF-3 fraction (>10,000 fold) from DNA affinity chromatography and in the crude extracts prepared from γ and α IFN treated HeLa cells [Fu et al.,*PROC. NATL. ACAD. SCI. USA,* 87 (1990)]. The antiserum was raised against a fusion protein [a cDNA fragment coding for part of 113 kD protein was inserted into pGEX-2T, a high expression vector in the *E. coli* [Smith et al., *PROC. NATL. ACAD. SCI. USA,* 83 (1986)] purified from *E. coli* [Smith et al., *GENE,* 67 (1988)]. The female NZW rabbits were immunized with 1 mg fusion protein in Freund's adjuvant. Two subsequent boosts two weeks apart were carried out with 500 mg fusion protein. The Western blot was carried out with conditions described previously [Pine et al., *MOL. CELL. BIOL.,* 10 (1990)].

FIG. **10**(B) presents the results of a mobility shift assay showing that the anti-113 antiserum affects the ISGF-3 shift complex. Preimmune serum or the 113 kD antiserum was added to shift reaction carried out as described [Fu et al. *PROC. NATL. ACAD. SCI. USA,* 87 (1990); Kessler et al.

GENES & DEV, 4, (1990)] at room temperature for 20 min. then one-third of reaction material was loaded onto a 5% polyacrylamide gel. In addition unlabeled probe was included in one reaction to show specificity of the gel shift complexes.

FIG. **11** shows the results of experiments investigating the IFN-α dependent phosphorylation of 113, 91 and 84 kD proteins. Protein samples from cells treated in various ways after 60 min. exposure to $^{32}PO_4^{-3}$ were precipitated with antiserum to 113 kD protein. Lane 1, no treatment of cells; Lane 2, cells treated 7 min. with IFN-α. By comparison with the marker proteins labeled 200, 97.5, 69 and 46 kD (kilo daltons), the PO_4^{-3} labeled proteins in the precipitate are seen to be 113 and 91 kD. Lane 3, cells treated with IFN-γ overnight (no phosphorylated proteins) and then (Lane 4) treated with IFN-α for 7 min. show heavier phosphorylation of 113, 91 and 84 kD.

FIG. **12** is a chromatogram depicting the identification of phosphoamino acid. Phosphate labeled protein of 113, 91 or 84 kD size was hydrolyzed and chromatographed to reveal newly labeled phosphotyrosine. Cells untreated with IFN showed only phosphoserine label. (P Ser=phosphoserine; P Thr=phosphotireonine; P Tyr=phosphotyrosine.

FIG. **13** depicts (A) the deduced amino acid sequence (SEQ ID NO:8) of and (B–C) the DNA sequence (SEQ ID NO:7) encoding the murine 91 kD intracellular receptor recognition factor.

FIG. **14** depicts (A) the deduced amino acid sequence (SEQ ID NO:10) of and (B–D) the DNA sequence (SEQ ID NO:9) encoding the 13sf1 intracellular receptor recognition factor.

FIG. **15** depicts (A) the deduced amino acid sequence (SEQ ID NO:12) of and (B–E) the DNA sequence (SEQ ID NO:11) encoding the 19sf6 intracellular receptor recognition factor.

FIG. **16**. Determination of molecular weights of Stat91 and phospho Stat91 by native gel analysis.

A) Western blot analysis of fractions from affinity purification. Extracts from human FS2 fibroblasts treated with IFN-γ (Ext), the unbound fraction (Flow), the fraction washed with Buffer AO.2 (AO.2), and the bound fraction eluted with buffer AO.8(AO.8) were immunoblotted with anti-91T.

B) Native gel analysis. Phosphorylated Stat91 (the AO.8 fraction from A) and unphosphorylated Stat91 (the Flow fraction from A) were analyzed on 4.5%, 5.5%, 6.5% and 7.5% native polyacrylamide gels followed by immunoblotting with anti-91T. The top of gels (TOP) and the migration position of bromophenol blue (BPB) are indicated.

C) Ferguson plots. The relative mobilities (Rm) of the Stat91 and phospho Stat91 were obtained from FIG. 1B (see Experimental Procedures). Closed circle: Chicken egg albumin (45 kD); Cross: Bovine serum albumin, monomer (66 kD); Open square: Bovine serum albumin, dimer (132 kD); Open circle: Urease, trimer (272 kD); Open triangle: Unphosphorylated Stat91; Closed triangle: Phosphorylated Stat91.

D) Determination of molecular weights from the standard curve. The molecular weights of phosphorylated and unphosphorylated Stat91 proteins (indicated as closed and open arrows, respectively) were obtained by extrapolation of their retardation coefficients.

FIG. **17**. Determination of molecular weights by glycerol gradients.

A) Western blot analysis. Extracts from human Bud8 fibroblasts treated with IFN-γ (the rightmost lane) and every other fraction from fraction 16 to 34 were analyzed on 7.5% SDS-PAGE followed by immunobloting with anti-91T. The peak of phosphorylated Stat91 (fraction 20) and the peak of unphosphorylated Stat91 (fraction 30) were indicated by a closed and open arrow, respectively.

B) Mobility shift analysis. Every other fractions from the gradients were analyzed.

C) Graphic representation of the data from A and B. Peak fraction numbers of protein standards are plotted versus their molecular weight. The position of peaks (of phosphorylated and unphosphorylated Stat91 protein are indicated by the closed and open arrows, respectively. Standards are ferritin (Fer, 440 kD), catalase (Cat, 232 kD), ferritin half unit (Fer ½, 220 kD)), aldolase (Ald, 158 kD), bovine serum albumin (BSA, 68 kD).

FIG. **18**. Stat91 in cell extracts binds DNA as a dimer.

A) Wester blot analysis. Extracts from stable cell lines expressing either Stat84 (C84), or Stat91L (C91L) or both (Cmx) were analyzed on 7.5% SDS-PAGE followed by immunobloting with anti-91.

B) Gel mobility shift analysis. Extracts from stable cell lines (FIG. 3A) untreated (–)or treated with IFN-γ(+) were analyzed. The positions of Stat91 homodimer (91L), Stat84 homodimer (84), and the heterodimer (84*91) are indicated.

FIG. **19**. Formation of herterodimer by denaturation and renaturation. Cytoplasmic (Left Panel) or nuclear extracts (Right Panel) from IFNγ-treated cell lines expressing either Stat84 (C84) or Stat91 (C91) were analyzed by gel mobility shift assays. +: with addition; –: without addition; D/R: samples were subjected to guanidinium hydrochloride denaturation and renaturation treatment.

FIG. **20**. Diagramatic representation of dissociation and reassociation analysis.

FIG. **21**. Dissociation-reassociation analysis with peptides. Gel mobility shift analysis with IFN-γ treated nuclear extracts from cell lines expressing Stat91L (C91L, lane 15) or Stat84 (C84, lane 14) or mixture of both (lane 1–13, 16–18) in the presence of increasing concentrations of various peptides. 91-Y, unphosphorylated peptide from Stat91 (LDGPKGTGYIKTELI) (SEQ. ID NO.:18); 91Y-p, phosphotyrosyl peptide from Stat91 (GY*IKTE) (SEQ ID NO.:19); 113Y-p, phosphotyrosyl peptide with high binding affinity to Src SH2 domain (EPQY*EEIPIYL, Songyang et al., 1993, Cell 72:767–778) (SEQ. ID NO.:21). Final concentrations of peptides added: 1 μM (lane 8), 4 μM (lane 2,5, 11), 10 μM (lane 9), 40 μM (lane 3, 6, 10, 12, 14–18), 160 μM (lane 4, 7, 13). +: with addition; –: without addition. Right panel: antiserum tests for identity of gel-shift bands (see FIG. **3**).

FIG. **22**. Dissociation-reassociation analysis with GST fusion proteins. A) SDS-PAGE (12%) analysis of purified GST fusion proteins as visualized by Commasie blue. GST-91 SH3, native SH2 domain of Stat91; GST-91 mSH2, R^{602} to L^{602} mutant; GST-91 SH3, SH3 domain of Stat91; GST Src SH2, the SH2 domain of src protein. Same amounts (1 μg) of each fusion proteins were loaded. Protein markers were run in lane 1 as indicated.

B) Dissociation-reassociation analysis similar to FIG. **6**. Dissociating agents were GST fusion proteins purified from bacterial expression as shown above. Final concentrations of fusion proteins added are 0.5 μM (lanes 2, 5, 8, 11, 14), 2.5 μM (lanes 3, 6, 9, 12, 15) and 5 μM (lanes 4, 7, 10, 13, 17, 18). +: with addition; –: without addition; FP: fusion proteins.

FIG. 23. Comparison of Stat91 SH2 structure with known SH2 structures. The Stat91 sequence is disclosed herein (SEQ ID NO:4). The structures used for the other SH2s are Src (Waksman et al., 1992, Nature 358:646–653) (SEQ ID NO:22), Abl (Overduin et al., 1992, Proc. Natl. Acad. Sci. USA 89:11673–77 and 1992, Cell 70:697–704) (SEQ ID NO:23, Lck (Eck et al., 1993, Nature 362:87–91) (SEQ ID NO:24), and p85αN (Booker et al., 1992, Nature 358:684–687) (SEQ ID NO:25). The alignment of the determined structures is by direct coordinate superimposition of the backbone structures. The names of secondary structural features and significant residues is based on the scheme of Eck et al., 1993. The boundaries and extents of the structure features are indicated by [- - -]. The starting numbers for the parent sequences are shown in parentheses. Experimentally determined structurally conserved regions are from Src, p85α, and Abl (Cowburn, unpublished). The root mean square deviation of three-dimensionally aligned structures differs by less than 1 Angstrom for the backbone non-hydrogen atoms in the sections marked by the XXX.

DETAILED DESCRIPTION

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual" (1982); "DNA Cloning: A Practical Approach," Volumes I and II (D. N. Glover ed. 1985); "Oligonucleotide Synthesis" (M. J. Gait ed. 1984); "Nucleic Acid Hybridization" [B. D. Hames & S. J. Higgins eds. (1985)]; "Transcription And Translation" [B. D. Hames & S. J. Higgins, eds. (1984)]; "Animal Cell Culture" [R. I. Freshney, ed. (1986)]; "Immobilized Cells And Enzymes" [IRL Press, (1986)]; B. Perbal, "A Practical Guide To Molecular Cloning" (1984).

Therefore, if appearing herein, the following terms shall have the definitions set out below.

The terms "receptor recognition factor", "receptor recognition-tyrosine kinase factor", "receptor recognition factor/tyrosine kinase substrate", "receptor recognition/transcription factor", "recognition factor" and "recognition factor protein(s)" and any variants not specifically listed, may be used herein interchangeably, and as used throughout the present application and claims refer to proteinaceous material including single or multiple proteins, and extends to those proteins having the amino acid sequence data described herein and presented in FIG. 1 (SEQ ID NO:2), FIG. 2 (SEQ ID NO:4) and in FIG. 3 (SEQ ID NO:6), and the profile of activities set forth herein and in the Claims. Accordingly, proteins displaying substantially equivalent or altered activity are likewise contemplated. These modifications may be deliberate, for example, such as modifications obtained through site-directed mutagenesis, or may be accidental, such as those obtained through mutations in hosts that are producers of the complex or its named subunits. Also, the terms "receptor recognition factor", "recognition factor" and "recognition factor protein(s)" are intended to include within their scope proteins specifically recited herein as well as all substantially homologous analogs and allelic variations.

The amino acid residues described herein are preferred to be in the "L" isomeric form. However, residues in the "D" isomeric form can be substituted for any L-amino acid residue, as long as the desired fuctional property of immunoglobulin-binding is retained by the polypeptide. NH2 refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxy terminus of a polypeptide. In keeping with standard polypeptide nomenclature, *J. Biol. Chem.*, 243:3552–59 (1969), abbreviations for amino acid residues are shown in the following Table of Correspondence:

TABLE OF CORRESPONDENCE

SYMBOL		
1-Letter	3-Letter	AMINO ACID
Y	Tyr	tyrosine
G	Gly	glycine
F	Phe	phenylalanine
M	Met	methionine
A	Ala	alanine
S	Ser	serine
I	Ile	isoleucine
L	Leu	leucine
T	Thr	threonine
V	Val	valine
P	Pro	proline
K	Lys	lysine
H	His	histidine
Q	Gln	glutamine
E	Glu	glutamic acid
W	Trp	tryptophan
R	Arg	arginine
D	Asp	aspartic acid
N	Asn	asparagine
C	Cys	cysteine

It should be noted that all amino-acid residue sequences are represented herein by formulae whose left and right orientation is in the conventional direction of amino-terminus to carboxy-terminus. Furthermore, it should be noted that a dash at the beginning or end of an amino acid residue sequence indicates a peptide bond to a further sequence of one or more amino-acid residues. The above Table is presented to correlate the three-letter and one-letter notations which may appear alternately herein.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication in vivo; i.e., capable of replication under its own control.

A "vector" is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in its either single stranded form, or a double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, inter alia, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having a sequence homologous to the mRNA).

An "origin of replication" refers to those DNA sequences that participate in DNA synthesis.

A DNA "coding sequence" is a double-stranded DNA sequence which is transcribed and translated into a polypeptide in vivo when placed under the control of appropriate

regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the −10 and −35 consensus sequences.

An "expression control sequence" is a DNA sequence that controls and regulates the transcription and translation of another DNA sequence. A coding sequence is "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then translated into the protein encoded by the coding sequence.

A "signal sequence" can be included before the coding sequence. This sequence encodes a signal peptide, N-terminal to the polypeptide, that communicates to the host cell to direct the polypeptide to the cell surface or secrete the polypeptide into the media, and this signal peptide is clipped off by the host cell before the protein leaves the cell. Signal sequences can be found associated with a variety of proteins native to prokaryotes and eukaryotes.

The term "oligonucleotide", as used herein in referring to the probe of the present invention, is defined as a molecule comprised of two or more ribonucleotides, preferably more than three. Its exact size will depend upon many factors which, in turn, depend upon the ultimate function and use of the oligonucleotide.

The term "primer" as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product, which is complementary to a nucleic acid strand, is induced, i.e., in the presence of nucleotides and an inducing agent such as a DNA polymerase and at a suitable temperature and pH. The primer may be either single-stranded or double-stranded and must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon many factors, including temperature, source of primer and use of the method. For example, for diagnostic applications, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15–25 or more nucleotides, although it may contain fewer nucleotides.

The primers herein are selected to be "substantially" complementary to different strands of a particular target DNA sequence. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to hybridize therewith and thereby form the template for the synthesis of the extension product.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

A cell has been "transformed" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

Two DNA sequences are "substantially homologous" when at least about 75% (preferably at least about 80%, and most preferably at least about 90 or 95%) of the nucleotides match over the defined length of the DNA sequences. Sequences that are substantially homologous can be identified by comparing the sequences using standard software available in sequence data banks, or in a Southern hybridization experiment under, for example, stringent conditions as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Maniatis et al., supra; DNA Cloning, Vols. I & II, supra; Nucleic Acid Hybridization, supra.

A "heterologous" region of the DNA construct is an identifiable segment of DNA within a larger DNA molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains introns, or synthetic sequences having codons different than the native gene). Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

An "antibody" is any immunoglobulin, including antibodies and fragments thereof, that binds a specific epitope. The term encompasses polyclonal, monoclonal, and chimeric antibodies, the last mentioned described in further detail in U.S. Pat. Nos. 4,816,397 and 4,816,567.

An "antibody combining site" is that structural portion of an antibody molecule comprised of heavy and light chain variable and hypervariable regions that specifically binds antigen.

The phrase "antibody molecule" in its various grammatical forms as used herein contemplates both an intact immunoglobulin molecule and an immunologically active portion of an immunoglobulin molecule.

Exemplary antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and those portions of an immunoglobulin molecule that contains the paratope, including those portions known in the art as Fab, Fab', $F(ab')_2$ and F(v), which portions are preferred for, use in the therapeutic methods described herein.

Fab and $F(ab')_2$ portions of antibody molecules are prepared by the proteolytic reaction of papain and pepsin, respectively, on substantially intact antibody molecules by methods that are well-known. See for example, U.S. Pat. No. 4,342,566 to Theofilopolous et al. Fab' antibody molecule portions are also well-known and are produced from $F(ab')_2$ portions followed by reduction of the disulfide bonds linking the two heavy chain portions as with mercaptoethanol, and followed by alkylation of the resulting protein mercaptan with a reagent such as iodoacetamide. An antibody containing intact antibody molecules is preferred herein.

The phrase "monoclonal antibody" in its various grammatical forms refers to an antibody having only one species of antibody combining site capable of immunoreacting with a particular antigen. A monoclonal antibody thus typically displays a single binding affinity for any antigen with which it immunoreacts. A monoclonal antibody may therefore contain an antibody molecule having a plurality of antibody combining sites, each immunospecific for a different antigen; e.g., a bispecific (chimeric) monoclonal antibody.

The phrase "pharmaceutically acceptable" refers to molecular entities and compositions that are physiologically tolerable and do not typically produce an allergic or similar untoward reaction, such as gastric upset, dizziness and the like, when administered to a human.

The phrase "therapeutically effective amount" is used herein to mean an amount sufficient to prevent, and preferably reduce by at least about 30 percent, more preferably by at least 50 percent, most preferably by at least 90 percent, a clinically significant change in the S phase activity of a target cellular mass, or other feature of pathology such as for example, elevated blood pressure, fever or white cell count as may attend its presence and activity.

A DNA sequence is "operatively linked" to an expression control sequence when the expression control sequence controls and regulates the transcription and translation of that DNA sequence. The term "operatively linked" includes having an appropriate start signal (e.g., ATG) in front of the DNA sequence to be expressed and maintaining the correct reading frame to permit expression of the DNA sequence under the control of the expression control sequence and production of the desired product encoded by the DNA sequence. If a gene that one desires to insert into a recombinant DNA molecule does not contain an appropriate start signal, such a start signal can be inserted in front of the gene.

The term "standard hybridization conditions" refers to salt and temperature conditions substantially equivalent to 5×SSC and 65° C. for both hybridization and wash.

In its primary aspect, the present invention concerns the identification of a receptor recognition factor, and the isolation and sequencing of a particular receptor recognition factor protein, that is believed to be present in cytoplasm and that serves as a signal transducer between a particular cellular receptor having bound thereto an equally specific polypeptide ligand, and the comparably specific transcription factor that enters the nucleus of the cell and interacts with a specific DNA binding site for the activation of the gene to promote the predetermined response to the particular polypeptide stimulus. The present disclosure confirms that specific and individual receptor recognition factors exist that correspond to known stimuli such as tumor necrosis factor, nerve growth factor, platelet-derived growth factor and the like. Specific evidence of this is set forth herein with respect to the interferons α and γ (IFNα and IFNγ).

A further property of the receptor recognition factors (also termed herein signal transducers and activators of transcription—STAT) is dimerization to form homodimers or heterodimers upon activation by phosphorylation of tyrosine. In a specific embodiment, infra, Stat91 and Stat84 form homodimers and a Stat91–Stat84 heterodimer. Accordingly, the present invention is directed to such dimers, which can form spontaneously by phophorylation of the STAT protein, or which can be prepared synthetically by chemically cross-linking two like or unlike STAT proteins.

The present receptor recognition factor is likewise noteworthy in that it appears not to be demonstrably affected by fluctuations in second messenger activity and concentration. The receptor recognition factor proteins appear to act as a substrate for tyrosine kinase domains, however do not appear to interact with G-proteins, and therefore do not appear to be second messengers.

A particular receptor recognition factor identified herein by SEQ ID NO:4, has been determined to be present in cytoplasm and serves as a signal transducer and a specific transcription factor in response to IFN-γ stimulation that enters the nucleus of the cell and interacts directly with a specific DNA binding site for the activation of the gene to promote the predetermined response to the particular polypeptide stimulus. This particular factor also acts as a translation protein and, in particular, as a DNA binding protein in response to interferon-γ stimulation. This factor is likewise noteworthy in that it has the following characteristics:

a) It interacts with an interferon-γ-bound receptor kinase complex;

b) It is a tyrosine kinase substrate; and

c) When phosphorylated, it serves as a DNA binding protein.

More particularly, the factor of SEQ ID NO:4 directly interacts with DNA after acquiring phosphate on tyrosine located at position 701 of the amino acid sequence. Also, interferon-γ-dependent activation of this factor occurs without new protein synthesis and appears within minutes of interferon-γ treatment, achieves maximum extent between 15 and 30 minutes thereafter, and then disappears after 2–3 hours.

In a particular embodiment, the present invention relates to all members of the herein disclosed family of receptor recognition factors except the 91 kD protein factors, specifically the proteins whose sequences are represented by one or more of SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8.

Subsequent to the filing of the initial applications directed to the present invention, the inventors have; termed each member of the family of receptor recognition factors as a signal transducer and activator of transcription (STAT) protein. Each STAT protein is designated by the apparent molecular weight (e.g., Stat113, Stat91, Stat84, etc.), or by the order in which it has been identified (e.g., Stat1α [Stat91], Stat1β [Stat84], Stat2 [Stat113], Stat3 [a murine protein described in U.S. application Ser. No. 08/126,588, filed Sep. 24, 1993 as 19sf6], and Stat4 [a murine STAT protein described in U.S. application Ser. No. 08/126,588,

filed Sep. 24, 1993 as 13sf1]). As will be readily appreciated by one of ordinary skill in the art, the choice of name has no effect on the intrinsic characteristics of the factors described herein, which were first disclosed in U.S. application Ser. No. 07/845,296, filed Mar. 19, 1992. The present inventors have chosen to adopt this newly derived terminology herein as a convenience to the skilled artisan who is familiar with the subsequently published papers relating to the same, and in accordance with the proposal to harmonize the terminology for the novel class of proteins, and nucleic acids encoding the proteins, disclosed by the instant inventors. The terms [molecular weight] kd receptor recognition factor, Stat[molecular weight], and Stat[number] are used herein interchangeably, and have the meanings given above. For example, the terms 91 kd protein, Stat91, and Stat1α refer to the same protein, and in the appropriate context refer to the nucleic acid molecule encoding such protein.

As stated above, the present invention also relates to a recombinant DNA molecule or cloned gene, or a degenerate variant thereof, which encodes a receptor recognition factor, or a fragment thereof, that possesses a molecular weight of about 113 kD and an amino acid sequence set forth in FIG. **1** (SEQ ID NO:2); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 113 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA sequence shown in FIG. **1** (SEQ ID NO:1). In another embodiment, the receptor recognition factor has a molecular weight of about 91 kD and the amino acid sequence set forth in FIG. **2** (SEQ ID NO:4) or FIG. **13** (SEQ ID NO:8); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 91 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqnece shown in FIG. **2** (SEQ ID NO:3) or FIG. **13** (SEQ ID NO:8). In yet a further embodiment, the receptor recognition factor has a molecular weight of about 84 kD and the amino acid sequence set forth in FIG. **3** (SEQ ID NO:6); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 84 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqnece shown in FIG. **3** (SEQ ID NO:5). In yet another embodiment, the receptor recognition factor has an amino acid sequence set forth in FIG. **14** (SEQ ID NO:10); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding such receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqnece shown in FIG. **14** (SEQ ID NO:9). In still another embodiment, the receptor recognition factor has an amino acid sequence set forth in FIG. **15** (SEQ ID NO:12); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding such receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqnece shown in FIG. **15** (SEQ ID NO:11).

The possibilities both diagnostic and therapeutic that are raised by the existence of the receptor recognition factor or factors, derive from the fact that the factors appear to participate in direct and causal protein-protein interaction between the receptor that is occupied by its ligand, and those factors that thereafter directly interface with the gene and effect transcription and accordingly gene activation. As suggested earlier and elaborated further on herein, the present invention contemplates pharmaceutical intervention in the cascade of reactions in which the receptor recognition factor is implicated, to modulate the activity initiated by the stimulus bound to the cellular receptor.

Thus, in instances where it is desired to reduce or inhibit the gene activity resulting from a particular stimulus or factor, an appropriate inhibitor of the receptor recognition factor could be introduced to block the interaction of the receptor recognition factor with those factors causally connected with gene activation. Correspondingly, instances where insufficient gene activation is taking place could be remedied by the introduction of additional quantities of the receptor recognition factor or its chemical or pharmaceutical cognates, analogs, fragments and the like.

As discussed earlier, the recognition factors or their binding partners or other ligands or agents exhibiting either mimicry or antagonism to the recognition factors or control over their production, may be prepared in pharmaceutical compositions, with a suitable carrier and at a strength effective for administration by various means to a patient experiencing an adverse medical condition associated specific transcriptional stimulation for the treatment thereof. A variety of administrative techniques may be utilized, among them parenteral techniques such as subcutaneous, intravenous and intraperitoneal injections, catheterizations and the like. Average quantities of the recognition factors or their subunits may vary and in particular should be based upon the recommendations and prescription of a qualified physician or veterinarian.

Also, antibodies including both polyclonal and monoclonal antibodies, and drugs that modulate the production or activity of the recognition factors and/or their subunits may possess certain diagnostic applications and may for example, be utilized for the purpose of detecting and/or measuring conditions such as viral infection or the like. For example, the recognition factor or its subunits may be used to produce both polyclonal and monoclonal antibodies to themselves in a variety of cellular media, by known techniques such as the hybridoma technique utilizing, for example, fused mouse spleen lymphocytes and myeloma cells. Likewise, small molecules that mimic or antagonize the activity(ies) of the receptor recognition factors of the invention may be discovered or synthesized, and may be used in diagnostic and/or therapeutic protocols.

The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal, antibody-producing cell lines can also be created by techniques other than fusion, such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al., "Hybridoma Techniques" (1980); Hammerling et al., "Monoclonal Antibodies And T-cell Hybridomas" (1981); Kennett et al., "Monoclonal Antibodies" (1980); see also U.S. Pat. Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,451,570; 4,466,917; 4,472,500; 4,491,632; 4,493,890.

Panels of monoclonal antibodies produced against recognition factor peptides can be screened for various properties; i.e., isotype, epitope, affinity, etc. Of particular interest are monoclonal antibodies that neutralize the activity of the recognition factor or its subunits. Such monoclonals can be readily identified in recognition factor activity assays. High affinity antibodies are also useful when immunoaffinity purification of native or recombinant recognition factor is possible.

Preferably, the anti-recognition factor antibody used in the diagnostic methods of this invention is an affinity purified polyclonal antibody. More preferably, the antibody is a monoclonal antibody (mAb). In addition, it is preferable for the anti-recognition factor antibody molecules used herein be in the form of Fab, Fab', $F(ab')_2$ or F(v) portions of whole antibody molecules.

As suggested earlier, the diagnostic method of the present invention comprises examining a cellular sample or medium by means of an assay including an effective amount of an antagonist to a receptor recognition factor/protein, such as an anti-recognition factor antibody, preferably an affinity-purified polyclonal antibody, and more preferably a mAb. In addition, it is preferable for the anti-recognition factor antibody molecules used herein be in the form of Fab, Fab', $F(ab')_2$ or F(v) portions or whole antibody molecules. As previously discussed, patients capable of benefiting from this method include those suffering from cancer, a precancerous lesion, a viral infection or other like pathological derangement. Methods for isolating the recognition factor and inducing anti-recognition factor antibodies and for determining and optimizing the ability of anti-recognition factor antibodies to assist in the examination of the target cells are all well-known in the art.

Methods for producing polyclonal anti-polypeptide antibodies are well-known in the art. See U.S. Pat. No. 4,493,795 to Nestor et al. A monoclonal antibody, typically containing Fab and/or $F(ab')_2$ portions of useful antibody molecules, can be prepared using the hybridoma technology described in *Antibodies—A Laboratory Manual*, Harlow and Lane, eds., Cold Spring Harbor Laboratory, New York (1988), which is incorporated herein by reference. Briefly, to form the hybridoma from which the monoclonal antibody composition is produced, a myeloma or other self-perpetuating cell line is fused with lymphocytes obtained from the spleen of a mammal hyperimmunized with a recognition factor-binding portion thereof, or recognition factor, or an origin-specific DNA-binding portion thereof.

Splenocytes are typically fused with myeloma cells using polyethylene glycol (PEG) 6000. Fused hybrids are selected by their sensitivity to HAT. Hybridomas producing a monoclonal antibody useful in practicing this invention are identified by their ability to immunoreact with the present recognition factor and their ability to inhibit specified transcriptional activity in target cells.

A monoclonal antibody useful in practicing the present invention can be produced by initiating a monoclonal hybridoma culture comprising a nutrient medium containing a hybridoma that secretes antibody molecules of the appropriate antigen specificity. The culture is maintained under conditions and for a time period sufficient for the hybridoma to secrete the antibody molecules into the medium. The antibody-containing medium is then collected. The antibody molecules can then be further isolated by well-known techniques.

Media useful for the preparation of these compositions are both well-known in the art and commercially available and include synthetic culture media, inbred mice and the like. An exemplary synthetic medium is Dulbecco's minimal essential medium (DMEM; Dulbecco et al., *Virol.* 8:396 (1959)) supplemented with 4.5 gm/l glucose, 20 mm glutamine, and 20% fetal calf serum. An exemplary inbred mouse strain is the Balb/c.

Methods for producing monoclonal anti-recognition factor antibodies are also well-known in the art. See Niman et al., *Proc. Natl. Acad. Sci. USA*, 80:4949–4953 (1983). Typically, the present recognition factor or a peptide analog is used either alone or conjugated to an immunogenic carrier, as the immunogen in the before described procedure for producing anti-recognition factor monoclonal antibodies. The hybridomas are screened for the ability to produce an antibody that immunoreacts with the recognition factor peptide analog and the present recognition factor.

The present invention further contemplates therapeutic compositions useful in practicing the therapeutic methods of this invention. A subject therapeutic composition includes, in admixture, a pharmaceutically acceptable excipient (carrier) and one or more of a receptor recognition factor, polypeptide analog thereof or fragment thereof, as described herein as an active ingredient. In a preferred embodiment, the composition comprises an antigen capable of modulating the specific binding of the present recognition factor within a target cell.

The preparation of therapeutic compositions which contain polypeptides, analogs or active fragments as active ingredients is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or suspensions, however, solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified. The active therapeutic ingredient is often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the composition can contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents which enhance the effectiveness of the active ingredient.

A polypeptide, analog or active fragment can be formulated into the therapeutic composition as neutralized pharmaceutically acceptable salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide or antibody molecule) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed from the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The therapeutic polypeptide-, analog- or active fragment-containing compositions are conventionally administered intravenously, as by injection of a unit dose, for example. The term "unit dose" when used in reference to a therapeutic composition of the present invention refers to physically discrete units suitable as unitary dosage for humans, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; i.e., carrier, or vehicle.

The compositions are administered in a manner compatible with the dosage formulation, and in a therapeutically effective amount. The quantity to be administered depends on the subject to be treated, capacity of the subject's immune system to utilize the active ingredient, and degree of inhibition or neutralization of recognition factor binding capacity desired. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner and are peculiar to each individual. However, suitable dosages may range from about 0.1 to 20, preferably about 0.5 to about 10, and more preferably one to several, milligrams of active ingredient per kilogram body weight of individual per day and depend on the route of administration. Suitable regimes for initial administration and booster shots are also variable, but are typified by an initial administration followed by repeated doses at one or more hour intervals by a subsequent injection or other administration. Alternatively, continuous intravenous infusion sufficient to maintain concentrations of ten nanomolar to ten micromolar in the blood are contemplated.

The therapeutic compositions may further include an effective amount of the factor/factor synthesis promoter

antagonist or analog thereof, and one or more of the following active ingredients: an antibiotic, a steroid. Exemplary formulations are given below:

Formulations

Ingredient	mg/ml
Intravenous Formulation I	
cefotaxime	250.0
receptor recognition factor	10.0
dextrose USP	45.0
sodium bisulfite USP	3.2
edetate disodium USP	0.1
water for injection q.s.a.d.	1.0 ml
Intravenous Formulation II	
ampicillin	250.0
receptor recognition factor	10.0
sodium bisulfite USP	3.2
disodium edetate USP	0.1
water for injection q.s.a.d.	1.0 ml
Intravenous Formulation III	
gentamicin (charged as sulfate)	40.0
receptor recognition factor	10.0
sodium bisulfite USP	3.2
disodium edetate USP	0.1
water for injection q.s.a.d.	1.0 ml
Intravenous Formulation IV	
recognition factor	10.0
dextrose USP	45.0
sodium bisulfite USP	3.2
edetate disodium USP	0.1
water for injection q.s.a.d.	1.0 ml
Intravenous Formulation V	
recognition factor antagonist	5.0
sodium bisulfite USP	3.2
disodium edetate USP	0.1
water for injection q.s.a.d.	1.0 ml

As used herein, "pg" means picogram, "ng" means nanogram, "ug" or "μg" mean microgram, "mg" means milligram, "ul" or "μl " mean microliter, "ml" means milliliter, "l " means liter.

Another feature of this invention is the expression of the DNA sequences disclosed herein. As is well known in the art, DNA sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host.

Such operative linking of a DNA sequence of this invention to an expression control sequence, of course, includes, if not already part of the DNA sequence, the provision of an initiation codon, ATG, in the correct reading frame upstream of the DNA sequence.

A wide variety of host/expression vector combinations may be employed in expressing the DNA sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and Synthetic DNA sequences. Suitable vectors include derivatives of SV40 and known bacterial plasmids, e.g., *E. coli* plasmids col E1, pCR1, pBR322, pMB9 and their derivatives, plasmids such as RP4; phage DNAS, e.g., the numerous derivatives of phage λ, e.g., NM989, and other phage DNA, e.g., M13 and Filamentous single stranded phage DNA; yeast plasmids such as the 2μ plasmid or derivatives thereof; vectors useful in eukaryotic cells, such as vectors useful in insect or mammalian cells; vectors derived from combinations of plasmids and phage DNAS, such as plasmids that have been modified to employ phage DNA or other expression control sequences; and the like.

Any of a wide variety of expression control sequences—sequences that control the expression of a DNA sequence operatively linked to it—may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include, for example, the early or late promoters of SV40, CMV, vaccinia, polyoma or adenovirus, the lac system, the trp system, the TAC system, the TRC system, the LTR system, the major operator and promoter regions of phage λ, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase (e.g., Pho5), the promoters of the yeast α-mating factors, and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof.

A wide variety of unicellular host cells are also useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of *E. coli*, Pseudomonas, Bacillus, Streptomyces, fungi such as yeasts, and animal cells, such as CHO, R1.1, B-W and L-M cells, African Green Monkey kidney cells (e.g., COS 1, COS 7, BSC1, BSC40, and BMT10), insect cells (e.g., Sf9), and human cells and plant cells in tissue culture.

It will be understood that not all vectors, expression control sequences and hosts will function equally well to express the DNA sequences of this invention. Neither will all hosts function equally well with the same expression system.

However, one skilled in the art will be able to select the proper vectors, expression control sequences, and hosts without undue experimentation to accomplish the desired expression without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must function in it. The vector's copy number, the ability to control that copy number, and the expression of any other proteins encoded by the vector, such as antibiotic markers, will also be considered.

In selecting an expression control sequence, a variety of factors will normally be considered. These include, for example, the relative strength of the system, its controllability, and its compatibility with the particular DNA sequence or gene to be expressed, particularly as regards potential secondary structures. Suitable unicellular hosts will be selected by consideration of, e.g., their compatibility with the chosen vector, their secretion characteristics, their ability to fold proteins correctly, and their fermentation requirements, as well as the toxicity to the host of the product encoded by the DNA sequences to be expressed, and the ease of purification of the expression products.

Considering these and other factors a person skilled in the art will be able to construct a variety of vector/expression control sequence/host combinations that will express the DNA sequences of this invention on fermentation or in large scale animal culture.

It is further intended that receptor recognition factor analogs may be prepared from nucleotide sequences of the protein complex/subunit derived within the scope of the present invention. Analogs, such as fragments, may be produced, for example, by pepsin digestion of receptor recognition factor material. Other analogs, such as muteins, can be produced by standard site-directed mutagenesis of receptor recognition factor coding sequences. Analogs exhibiting "receptor recognition factor activity" such as

small molecules, whether functioning as promoters or inhibitors, may be identified by known in vivo and/or in vitro assays.

As mentioned above, a DNA sequence encoding receptor recognition factor can be prepared synthetically rather than cloned. The DNA sequence can be designed with the appropriate codons for the receptor recognition factor amino acid sequence. In general, one will select preferred codons for the intended host if the sequence will be used for expression. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge, *Nature,* 292:756 (1981); Nambair et al., *Science,* 223:1299 (1984); Jay et al., *J. Biol. Chem.,* 259:6311 (1984).

Synthetic DNA sequences allow convenient construction of genes which will express receptor recognition factor analogs or "muteins". Alternatively, DNA encoding muteins can be made by site-directed mutagenesis of native receptor recognition factor genes or cDNAs, and muteins can be made directly using conventional polypeptide synthesis.

A general method for site-specific incorporation of unnatural amino acids into proteins is described in Christopher J. Noren, Spencer J. Anthony-Cahill, Michael C. Griffith, Peter G. Schultz, *Science,* 244:182–188 (April 1989). This method may be used to create analogs with unnatural amino acids.

The present invention extends to the preparation of antisense nucleotides and ribozymes that may be used to interfere with the expression of the receptor recognition proteins at the translational level. This approach utilizes antisense nucleic acid and ribozymes to block translation of a specific mRNA, either by masking that mRNA with an antisense nucleic. acid or cleaving it with a ribozyme.

Antisense nucleic acids are DNA or RNA molecules that are complementary to at least a portion of a specific mRNA molecule. (See Weintraub, 1990; Marcus-Sekura, 1988.) In the cell, they hybridize to that mRNA, forming a double stranded molecule. The cell does not translate an mRNA in this double-stranded form. Therefore, antisense nucleic acids interfere with the expression of mRNA into protein. Oligomers of about fifteen nucleotides and molecules that hybridize to the AUG initiation codon will be particularly efficient, since they are easy to synthesize and are likely to pose fewer problems than larger molecules when introducing them into receptor recognition factor-producing cells. Antisense methods have been used to inhibit the expression of many genes in vitro (Marcus-Sekura, 1988; Hambor et al., 1988).

Ribozymes are RNA molecules possessing the ability to specifically cleave other single stranded RNA molecules in a manner somewhat an alogous to DNA restriction endonucleases. Ribozymes were discovered from the observation that certain mRNAs have the ability to excise their own introns. By modifying the nucleotide sequence of these RNAs, researchers have been able to engineer molecules that recognize specific nucleotide sequences in an RNA molecule and cleave it (Cech, 1988.). Because they are sequence-specific, only mRNAs with particular sequences are inactivated.

Investigators have identified two types of ribozymes, Tetrahymena-type and "hammerhead"-type. (Hasselhoff and Gerlach, 1988) Tetrahymena-type ribozymes recognize four-base sequences, while "hammerhead"-type recognize eleven- to eighteen-base sequences. The longer the recognition sequence the more likely it is to occur exclusively in the target mRNA species. Therefore, hammerhead-type ribozymes are preferable to Tetrahymena-type ribozymes for inactivating a specific mRNA species, and eighteen base recognition sequences are preferable to shorter recognition sequences.

The DNA sequences described herein may thus be used to prepare antisense molecules against, and ribozymes that cleave mRNAs for receptor recognition factor proteins and their ligands.

The present invention also relates to a variety of diagnostic applications, including methods for detecting the presence of stimuli such as the earlier referenced polypeptide ligands, by reference to their ability to elicit the activities which are mediated by the present receptor recognition factor. As mentioned earlier, the receptor recognition factor can be used to produce antibodies to itself by a variety of known techniques, and such antibodies could then be isolated and utilized as in tests for the presence of particular transcriptional activity in suspect target cells.

As described in detail above, antibody(ies) to the receptor recognition factor can be produced and isolated by standard methods including the well known hybridoma techniques. For convenience, the antibody(ies) to the receptor recognition factor will be referred to herein as Ab_1 and antibody(ies) raised in another species as Ab_2.

The presence of receptor recognition factor in cells can be ascertained by the usual immunological procedures applicable to such determinations. A number of useful procedures are known. Three such procedures which are especially useful utilize either the receptor recognition factor labeled with a detectable label, antibody Ab_1 labeled with a detectable label, or antibody Ab_2 labeled with a detectable label. The procedures may be summarized by the following equations wherein the asterisk indicates that the particle is labeled, and "RRF" stands for the receptor recognition factor:

$$RRF^*+Ab_1=RRF^*Ab_1 \quad \text{A.}$$

$$RRF+Ab^*=RRFAb_1^* \quad \text{B.}$$

$$RRF+Ab_1+Ab_2^*=RRFAb_1Ab_2^* \quad \text{C.}$$

The procedures and their application are all familiar to those skilled in the art and accordingly may be utilized within the scope of the present invention. The "competitive" procedure, Procedure A, is described in U.S. Pat. Nos. 3,654,090 and 3,850,752. Procedure C, the "sandwich" procedure, is described in U.S. Pat. Nos. RE 31,006 and 4,016,043. Still other procedures are known such as the "double antibody", or "DASP" procedure.

In each instance, the receptor recognition factor forms complexes with one or more antibody(ies) or binding partners and one member of the complex is labeled with a detectable label. The fact that a complex has formed and, if desired, the amount thereof, can be determined by known methods applicable to the detection of labels.

It will be seen from the above, that a characteristic property of Ab_2 is that it will react with Ab_1. This is because Ab_1 raised in one mammalian species has been used in another species as an antigen to raise the antibody Ab_2. For example, Ab_2 may be raised in goats using rabbit antibodies as antigens. Ab_2 therefore would be anti-rabbit antibody raised in goats. For purposes of this description and claims, Ab_1 will be referred to as a primary or anti-receptor recognition factor antibody, and Ab_2 will be referred to as a secondary or anti-Ab_1 antibody.

The labels most commonly employed for these studies are radioactive elements, enzymes, chemicals which fluoresce when exposed to ultraviolet light, and others.

A number of fluorescent materials are known and can be utilized as labels. These include, for example, fluorescein, rhodamine and auramine. A particular detecting material is anti-rabbit antibody prepared in goats and conjugated with fluorescein through an isothiocyanate.

The receptor recognition factor or its binding partner(s) can also be labeled with a radioactive element or with an enzyme. The radioactive label can be detected by any of the currently available counting procedures. The preferred isotope may be selected from ^{3}H, ^{14}C, ^{32}P, ^{35}S, ^{36}Cl, ^{51}Cr, ^{57}Co, ^{58}Co, ^{59}Fe, ^{90}Y, ^{125}I, ^{131}I, and ^{186}Re.

Enzyme labels are likewise useful, and can be detected by any of the presently utilized colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques. The enzyme is conjugated to the selected particle by reaction with bridging molecules such as carbodiimides, diisocyanates, glutaraldehyde and the like. Many enzymes which can be used in these procedures are known and can be utilized. The preferred are peroxidase, β-glucuronidase, β-D-glucosidase, β-D-galactosidase, urease, glucose oxidase plus peroxidase and alkaline phosphatase. U.S. Pat. Nos. 3,654,090; 3,850,752; and 4,016,043 are referred to by way of example for their disclosure of alternate labeling material and methods.

A particular assay system developed and utilized in accordance with the present invention, is known as a receptor assay. In a receptor assay, the material to be assayed is appropriately labeled and then certain cellular test colonies are inoculated with a quantity of both the labeled and unlabeled material after which binding studies are conducted to determine the extent to which the labeled material binds to the cell receptors. In this way, differences in affinity between materials can be ascertained.

Accordingly, a purified quantity of the receptor recognition factor may be radiolabeled and combined, for example, with antibodies or other inhibitors thereto, after which binding studies would be carried out. Solutions would then be prepared that contain various quantities of labeled and unlabeled uncombined receptor recognition factor, and cell samples would then be inoculated and thereafter incubated. The resulting cell monolayers are then washed, solubilized and then counted in a gamma counter for a length of time sufficient to yield a standard error of <5%. These data are then subjected to Scatchard analysis after which observations and conclusions regarding material activity can be drawn. While the foregoing is exemplary, it illustrates the manner in which a receptor assay may be performed and utilized, in the instance where the cellular binding ability of the assayed material may serve as a distinguishing characteristic.

An assay useful and contemplated in accordance with the present invention is known as a "cis/trans" assay. Briefly, this assay employs two genetic constructs, one of which is typically a plasmid that continually expresses a particular receptor of interest when transfected into an appropriate cell line, and the second of which is a plasmid that expresses a reporter such as luciferase, under the control of a receptor/ligand complex. Thus, for example, if it is desired to evaluate a compound as a ligand for a particular receptor, one of the plasmids would be a construct that results in expression of the receptor in the chosen cell line, while the second plasmid would possess a promoter linked to the luciferase gene in which the response element to the particular receptor is inserted. If the compound under test is an agonist for the receptor, the ligand will complex with the receptor, and the resulting complex will bind the response element and initiate transcription of the luciferase gene. The resulting chemiluminescence is then measured photometrically, and dose response curves are obtained and compared to those of known ligands. The foregoing protocol is described in detail in U.S. Pat. No. 4,981,784 and PCT International Publication No. WO 88/03168, for which purpose the artisan is referred.

In a further embodiment of this invention, commercial test kits suitable for use by a medical specialist may be prepared to determine the presence or absence of predetermined transcriptional activity or predetermined transcriptional activity capability in suspected target cells. In accordance with the testing techniques discussed above, one class of such kits will contain at least the labeled receptor recognition factor or its binding partner, for instance an antibody specific thereto, and directions, of course, depending upon the method selected, e.g., "competitive", "sandwich", "DASP" and the like. The kits may also contain peripheral reagents such as buffers, stabilizers, etc.

Accordingly, a test kit may be prepared for the demonstration of the presence or capability of cells for predetermined transcriptional activity, comprising:

(a) a predetermined amount of at least one labeled immunochemically reactive component obtained by the direct or indirect attachment of the present receptor recognition factor or a specific binding partner thereto, to a detectable label;

(b) other reagents; and

(c) directions for use of said kit.

More specifically, the diagnostic test kit may comprise:

(a) a known amount of the receptor recognition factor as described above (or a binding partner) generally bound to a solid phase to form an immunosorbent, or in the alternative, bound to a suitable tag, or plural such end products, etc. (or their binding partners) one of each;

(b) if necessary, other reagents; and

(c) directions for use of said test kit.

In a further variation, the test kit may be prepared and used for the purposes stated above, which operates according to a predetermined protocol (e.g. "competitive", "sandwich", "double antibody", etc.), and comprises:

(a) a labeled component which has been obtained by coupling the receptor recognition factor to a detectable label;

(b) one or more additional immunochemical reagents of which at least one reagent is a ligand or an immobilized ligand, which ligand is selected from the group consisting of:

(i) a ligand capable of binding with the labeled component (a);

(ii) a ligand capable of binding with a binding partner of the labeled component (a);

(iii) a ligand capable of binding with at least one of the component(s) to be determined; and

(iv) a ligand capable of binding with at least one of the binding partners of at least one of the component(s) to be determined; and

(c) directions for the performance of a protocol for the detection and/or determination of one or more components of an immunochemical reaction between the receptor recognition factor and a specific binding partner thereto.

In accordance with the above, an assay system for screening potential drugs effective to modulate the activity of the receptor recognition factor may be prepared. The receptor recognition factor may be introduced into a test system, and the prospective drug may also be introduced into the result-

ing cell culture, and the culture thereafter examined to observe any changes in the transcriptional activity of the cells, due either to the addition of the prospective drug alone, or due to the effect of added quantities of the known receptor recognition factor.

PRELIMINARY CONSIDERATIONS

As mentioned earlier, the observation and conclusion underlying the present invention were crystallized from a consideration of the results of certain investigations with particular stimuli. Particularly, the present disclosure is illustrated by the results of work on protein factors that govern transcriptional control of IFNα-stimulated genes, as well as more recent data on the regulation of transcription of genes stimulated by IFNγ. The following is a brief discussion of the role that IFN is believed to play in the stimulation of transcription taken from Darnell et al. *THE NEW BIOLOGIST,* 2(10), (1990).

Activation of genes by IFNα occurs within minutes of exposure of cells to this factor (Larner et al., 1984, 1986) and is strictly dependent on the IFNα binding to its receptor, a 49-kD plasma membrane polypeptide (Uze et al., 1990). However, changes in intracellular second messenger concentrations secondary to the use of phorbol esters, calcium ionophores, or cyclic nucleotide analogs neither triggers nor blocks IFNα-dependent gene activation (Larner et al., 1984; Lew et al., 1989). No other polypeptide, even IFNγ, induces the set of interferon-stimulated genes (ISGs) specifically induced by IFNα. In addition, it has been found that IFNγ-dependent transcriptional stimulation of at least one gene in HeLa cells and in fibroblasts is also strictly dependent on receptor-ligand interaction and is not activated by induced changes in second messengers (Decker et al., 1989; Lew et al., 1989). These highly specific receptor-ligand interactions, as well as the precise transcriptional response, require the intracellular recognition of receptor occupation and the communication to the nucleus to be equally specific.

The activation of ISGs by IFNα is carried out by transcriptional factor ISGF-3, or interferon stimulated gene factor 3. This factor is activated promptly after IFNα treatment without protein synthesis, as is transcription itself (Larner et al., 1986; Levy et al., 1988; Levy et al., 1989). ISGF-3 binds to the ISRE, the interferon-stimulated response element, in DNA of the response genes (Reich et al., 1987; Levy et al., 1988), and this binding is affected by all of an extensive set of mutations that also affects the transcriptional function of the ISRE (Kessler et al., 1988a). Partially purified ISGF-3 containing no other DNA-binding components can stimulate ISRE-dependent in vitro transcription (Fu et al., 1990). IFN-dependent stimulation of ISGs occurs in a cycle, reaching a peak of 2 hours and declining promptly thereafter (Larner et al., 1986). ISGF-3 follows the same cycle (Levy et al., 1988, 1989). Finally, the presence or absence or ISGF3 in a variety of IFN-sensitive and IFN-resistant cells correlates with the transcription of ISGs in these cells (Kessler et al., 1988b).

ISGF-3 is composed of two subfractions, ISGF-3α and ISGF-3γ, that are found in the cytoplasm before IFN binds to its receptor (Levy et al., 1989). When cells are treated with IFNα, ISGF-3 can be detected in the cytoplasm within a minute, that is, some 3 to 4 minutes before any ISGF-3 is found in the nucleus (Levy et al., 1989). The cytoplasmic component ISGF-3γ can be increased in HeLa cells by pretreatment with IFNγ, but IFNγ does not by itself activate transcription of ISGs nor raise the concentration of the complete factor, ISGF-3 (Levy et al., 1990). The cytoplasmic localization of the proteins that interact to constitute ISGF-3 was proved by two kinds of experiments. When cytoplasm of IFNγ-treated cells that lack ISGF-3 was mixed with cytoplasm of IFNα-treated cells, large amounts of ISGF-3 were formed (Levy et al., 1989). (It was this experiment that indicated the existence of an ISGF-3γ component and an ISGF-3α component of ISGF-3).

In addition, Dale et al. (1989) showed that enucleated cells could respond to IFNα by forming a DNA-binding protein that is probably the same as ISGF-3.

The ISGF-3γ component is a 48-kD protein that specifically recognizes the ISRE (Kessler et al., 1990; Fu et al., 1990). Three other proteins, presumably constituting the ISGF-3α component, were found in an ISGF-3 DNA complex (Fu et al., 1990). The entirety of roles of, or the relationships among these three proteins are not yet known, but it is clear that ISGF-3 is a multimeric protein complex. Since the binding of IFNα to the cell surface converts ISGF-3α from an inactive to an active status within a minute, at least one of the proteins constituting ISGF-3α must be affected promptly, perhaps by a direct interaction with the IFNα receptor.

The details of how the ISGF-3γ component and the three other proteins are activated by cytoplasmic events and then enter the nucleus to bind the ISRE and increase transcription are not entirely known. Further studies of the individual proteins, for example, with antibodies, are presented herein. For example, it is clear that, within 10 minutes of IFNα treatment, there is more ISGF-3 in the nucleus than in the cytoplasm and that the complete factor has a much higher affinity for the ISRE than the 48-kD ISGF-3γ component by itself (Kessler et al., 1990).

In summary, the attachment of interferon-α (IFN-α) to its specific cell surface receptor activates the transcription of a limited set of genes, termed ISGs for "interferon stimulated genes" [Larner et al., *PROC. NATL. ACAD. SCI. USA,* 81 (1984); Larner et al., *J. BIOL. CHEM.,* 261 (1986); Friedman et al., *CELL,* 38 (1984)]). The observation that agents that affect second messenger levels do not activate transcription of these genes, led to the proposal that protein:protein interactions in the cytoplasm beginning at the IFN receptor might act directly in transmitting to the nucleus the signal generated by receptor occupation [Levy et al., *NEW BIOLOGIST,* 2 (1991)].

To test this hypothesis, the present applicants began experiments in the nucleus at the activated genes. Initially, the ISRE and ISGF-3 were discovered [Levy et al., *GENES & DEV.,* 2 (1988)].

Partial purification of ISGF-3 followed by recovery of the purified proteins from a specific DNA-protein complex revealed that the complete complex was made up of four proteins [Fu et al., *PROC. NATL. ACAD. SCI. USA,* 87 (1990); Kessler et al., *GENES & DEV.,* 4 (1990)]. A 48 kD protein termed ISGF-3γ, because pre-treatment of HeLa cells with IFN-γ increased its presence, binds DNA weakly on its own [Ibid.; and Levy et al., *THE EMBO. J.,* 9 (1990)]. In combination with the IFN-α activated proteins, termed collectively the ISGF-3α proteins, the ISGF-3γ forms a complex that binds the ISRE with a 50-fold higher affinity [Kessler et al., *GENES & DEV.,* 4 (1990)]. The ISGF-3α proteins comprise a set of polypeptides of 113, 91 and 84 kD. All of the ISGF-3 components initially reside in the cell cytoplasm [Levy et al., *GENES & DEV.,* 3 (1989); Dale et al., *PROC. NATL. ACAD. SCI. USA,* 86 (1989)]. However after only about five minutes of IFN-α treatment the active complex is found in the cell nucleus, thus confirming these

proteins as a possible specific link from an occupied receptor to a limited set of genes [Levy et al., *GENES & DEV.*, 3 (1989)].

In accordance with the present invention, specific proteins comprising receptor recognition factors have been isolated and sequenced. These proteins, their fragments, antibodies and other constructs and uses thereof, are contemplated and presented herein. To understand the mechanism of cytoplasmic activation of the ISGF-3α proteins as well as their transport to the nucleus and interaction with ISGF-3γ, this factor has been purified in sufficient quantity to obtain peptide sequence from each protein. Degenerate deoxyoligonucleotides that would encode the peptides were constructed and used in a combination of cDNA library screening and PCR amplification of cDNA products copied from mRNA to identify cDNA clones encoding each of the four proteins. What follows in the examples presented herein a description of the final protein preparations that allowed the cloning of cDNAs encoding all the proteins, and the primary sequence of the 113 kD protein arising from a first gene, and the primary sequences of the 91 and 84 kD proteins which appear to arise from two differently processed RNA products from another gene. Antisera against portions of the 84 and 91 kD proteins have also been prepared and bind specifically to the ISGF-3 DNA binding factor (detected by the electrophoretic mobility shift assay with cell extracts) indicating that these cloned proteins are indeed part of ISGF-3. The availability of the cDNA and the proteins they encode provides the necessary material to understand how the liganded IFN-α receptor causes immediate cytoplasmic activation of the ISGF-3 protein complex, as well as to understand the mechanisms of action of the receptor recognition factors contemplated herein. The cloning of each of ISGF3-α proteins, and the evaluation and confirmation of the particular role played by the 91 kD protein as a messenger and DNA binding protein in response to IFN-γ activation, including the development and testing of antibodies to the receptor recognition factors of the present invention, are all presented in the examples that follow below.

EXAMPLE 1

To purify relatively large amounts of ISGF-3, HeLa cell nuclear extracts were prepared from cells treated overnight (16–18 h) with 0.5 ng/ml of IFN-γ and 45 min. with IFN-α (500 u/ml). The steps used in the large scale purification were modified slightly from those described earlier in the identification of the four ISGF-3 proteins.

Accordingly, nuclear extracts were made from superinduced HeLa cells [Levy et al., *THE EMBO. J.*, 9 (1990)] and chromatographed as previously described [Fu et al., *PROC. NATL. ACAD. SCI. USA*, 87 (1990)] on: phosphocellulose P-11, heparin agarose (Sigma); DNA cellulose (Boehringer Mannheim; flow through was collected after the material was adjusted to 0.28M KCl and 0.5% NP-40); two successive rounds of ISRE oligo affinity column (1.8 ml column, eluted with a linear gradient of 0.05 to 1.0M KCl); a point mutant ISRE oligonucleotide affinity column (flow through was collected after the material was adjusted to 0.28M KCl); and a final round on the ISRE oligonucleotide column (material was eluted in a linear 0.05 to 1.0M NaCl gradient adjusted to 0.05% NP-40). Column fractions containing ISGF-3 were subsequently examined for purity by SDS PAGE/silver staining and pooled appropriately. The pooled fractions were concentrated by a centricon-10 (Amicon). The pools of fractions from preparations 1 and 2 were combined and run on a 10 cm wide, 1.5 mm thick 7.5% SDS polyacrylamide gel. The proteins were electroblotted to nitrocellulose for 12 hrs at 20 volts in 12.5% MeOH, 25 mM Tris, 190 mM glycine. The membrane was stained with 0.1% Ponceau Red (in 1% acetic acid) and the bands of 113 kD, 91 kD, 84 kD, and 48 kD excised and subjected to peptide analysis after tryptic digestion [Wedrychowski et al., *J. BIOL. CHEM.*, 265 (1990); Aebersold et al., *PROC. NATL. ACAD. SCI. USA*, 84 (1987)]. The resulting peptide sequences for the 91 kD and 84 kD proteins are indicated in FIG. **6**. Degenerate oligonucleotides were designed based on the peptide sequences t19, t13b and t27: (Forward and Reverse complements are denoted by F and R:

```
19F   AACGTIGACCAATTNAACATG      (SEQ ID NO:14)
        T     T GC    T
          T
13bR GTCGATGTTNGGGTANAG         (SEQ ID NO:15)
     A A A    A     A
27R  GTACAAITCAACCAGNGCAA       (SEQ ID NO:16)
       T    TG T    T
```

The final ISRE oligonucleotide affinity selection yielded material with the SDS polyacrylamide gel electrophoretic pattern shown in FIG. **4** (left). This gel represented about 1.5% of the available material purified from over 200 L of appropriately treated HeLa cells. While 113, 91, 84 and 48 kD bands were clearly prominent in the final purified preparation (see FIG. **4**, right panel), there were also two prominent contaminants of about 118 and 70 kD and a few of other contaminants in lower amounts. [Amino acid sequence data have shown that the contaminants of 86 kD and 70 kD are the KU antigen, a widely-distributed protein that binds DNA termini. However in the specific ISGF-3: ISRE complex there is no KU antigen and therefore it has been assigned no role in IFN-dependent transcriptional stimulation, [Wedrychowski et al., *J. BIOL. CHEM.*, 265 (1990)]].

Since the mobility of the 113, 91, 84, and 48 kD proteins could be accurately marked by comparison with the partially purified proteins characterized in previous experiments [Fu et al., *PROC. NATL. ACAD. SCI. USA*, 87 (1990)], further purification was not attempted at this stage. The total purified sample from 200 L of HeLa cells was loaded onto one gel, subjected to electrophoresis, transferred to nitrocellulose and stained with Ponceau red. The 113, 84, 91, and 48 kD protein bands were separately excised and subjected to peptide analysis as described [Aebersold et al., *PROC. NATL. ACAD. SCI. USA*, 84 (1987)]. Released peptides were collected, separated by HPLC and analyzed for sequence content by automated Edman degradation analysis.

Accordingly, the use of the peptide sequence data for three of four peptides from the 91 kD protein and a single peptide derived from the 84 kD protein is described herein. The peptide sequence and the oligonucleotides constructed from them are given in the legend to FIG. **4** or **6**. When oligonucleotides 19F and 13bR were used to prime synthesis from a HeLa cell cDNA library, a PCR product of 475 bp was generated. When this product was cloned and sequenced it encoded the 13a peptide internally. Oligonucleotide 27R derived from the only available 84 kD peptide sequence was used in an anchored PCR procedure amplifying a 405 bp segment of DNA. This 405 bp amplified sequence was identical to an already sequenced region of the 91 kD protein. It was then realized that the peptide t27 sequence was contained within peptide t19 and that the 91 and 84 kD proteins must be related (see FIGS. **5** & **7**). Oligonucleotides 19F and 13a were also used to select candidate cDNA clones from a cDNA library made from mRNA prepared after 16 hr. of IFN-γ and 45 min. of IFN-α treatment.

Of the numerous cDNA clones that hybridized these oligonucleotides and also the cloned PCR products, one cDNA clone, E4, contained the largest open reading frame flanked by inframe stop codons. Sequence of peptides t19, t13a, and t13b were contained in this 2217 bp ORF (see FIG. 6) which was sufficient to encode a protein of 739 amino acids (calculated molecular weight of 86 kD). The codon for the indicated initial methionine was preceded by three in frame stop codons. This coding capacity has been confirmed by translating in vitro an RNA copy of the E4 clone yielding product of nominal size of 86 kD, somewhat shorter than the in vitro purified 91 kD protein discussed earlier (data not shown). Perhaps this result indicates post-translational modification of the protein in the cell.

A second class of clones was also identified (see FIG. 5). E3, the prototype of this class was identical to E4 from the 5' end to bp 2286 (aa 701) at which point the sequences diverged completely. Both cDNAs terminated with a poly (A) tail. Primer extension analysis suggested another ~150 bp were missing from the 5' end of both mRNAs. DNA probes were made from the clones representing both common and unique sequences for use in Northern blot analyses. The preparation of the probes is as follows: 20 mg of cytoplasmic RNA (0.5% NP-40 lysate) of IFN-α treated (6 h) HeLa RNA was fractionated in a 1% agarose, 6% formaldehyde gel (in 20 mM MOPS, 5 mM NaAc, 1 mM EDTA, pH 7.0) for 4.5 h at 125 volts. The RNA was transferred in 20×SSC to Hybond-N (Amersham), UV crosslinked and hybridized with 1×106 cpm/ml of the indicated probes (1.5×10^8 cpm/mg).

Probes from regions common to E3 and E4 hybridized to two RNA species of approximately 3.1 KB and 4.4 KB. Several probes derived from the 3' non-coding end of, E4, which were unique to E4, hybridized only the larger RNA species. A labeled DNA probe from the unique 3' non-coding end of E3 hybridized only the smaller RNA species.

Review of the sequence at the site of 3' discontinuity between E3 and E4 suggested that the shorter mRNA results from choice of a different poly(A) site and 3' exon that begins at bp 2286 (the calculated molecular weight from the E3. The last two nucleotides before the change are GT followed by GT in E3 in line with the consensus nucleotides at an exon-intron junction. Since the ORF of E4 extends to bp 2401 it encodes a protein that is 38 amino acids longer than the one encoded by E3, but is otherwise identical (ORF is 82 kD).

Since there is no direct assay for the activity of the 91 or 84 kD protein, an independent method was needed to determine whether the cDNA clones we had isolated did indeed encode proteins that are part of ISGF-3. For this purpose antibodies were initially raised against the sequence from amino acid 597 to amino acid 703 (see FIG. 6) by expressing this peptide in the pGEX-3X vector (15) as a bacterial fusion protein. This antiserum (a42) specifically recognized the 91 kD and 84 kD proteins in both crude extracts and purified ISGF-3 (see FIG. 7*a*). More importantly this antiserum specifically affected the ISGF-3 band in a mobility shift assay using the labeled ISRE oligonucleotide (see FIG. 7*b*) confirming that the isolated 91 kD and 84 kD cDNA clones (E4 and E3) represent a component of ISGF-3. Additional antisera were raised against the amino terminus and carboxy terminus of the protein encoded by E4. The amino terminal 59 amino acids that are common to both proteins and the unique carboxy terminal 34 amino acids encoded only by the larger mRNA were expressed as fusion proteins in pGEX-3X for immunization of rabbits. Western blot analysis with highly purified ISGF-3 demonstrated that the amino terminal antibody (a55) recognized both the 91 kD and 84 kD proteins as expected. However, the other antibody (a57) recognized only the 91 kD protein confirming our assumption that the larger mRNA (4.4 KB) and larger cDNA encodes the 91 kD protein while the shorter mRNA (3.1 KB) and cDNA encodes the 84 kD protein (see FIG. 7*a*).

EXAMPLE 2

In this example, the cloning of the 113 kD protein that comprises one of the three ISGF-3α components is disclosed.

From SDS gels of highly purified ISGF-3, the 113 kD band was identified, excised and subjected to cleavage and peptide sequence analysis [Aebersold et al., *PROC. NATL. ACAD. SCI. USA,* 87 (1987)]. Five peptide sequences (A–E) were obtained (FIG. 8A). Degenerate oligonucleotide probes were designed according to these peptides which then were radiolabeled to search a human cDNA library for clones that might encode the 113 kD protein. Eighteen positive cDNA clones were recovered from 2.5×10^5 phage plaques with the probe derived from peptide E (FIG. 8A, and the legend). Two of them were completely sequenced. Clone f11 contained a 3.2 KB cDNA, and clone ka31 a 2.6 KB cDNA that overlapped about 2 KB but which had a further extended 5' end in which a candidate AUG initiation codon was found associated with a well-conserved Kozak sequence [Kozak, *NUCLEIC ACIDS RES.,* 12 (1984)].

In addition to the phage cDNA clones, a PCR product made between oligonucleotides that encoded peptide D and E also yielded a 474 NT fragment that when sequenced was identical with the cDNA clone in this region. A combination of these clones f11 and ka31 revealed an open reading frame capable of encoding a polypeptide of 851 amino acids (FIG. 8A). These two clones were joined within their overlapping region and RNA transcribed from this recombinant clone was translated in vitro yielding a polypeptide that migrated in an SDS gel with a nominal molecular weight of 105 kD (FIG. 9A). An appropriate clone encoding the 91 kD protein was also transcribed and the RNA translated in the same experiment. Since both the apparently complete cDNA clones for the 113 kD protein and the 91 kD protein produce RNAs that when translated into proteins migrate somewhat faster than the proteins purified as ISGF-3 components, it is possible that the proteins undergo post-translational modification in the cell causing them to be slightly retarded during electrophoresis. When a 660 bp cDNA encoding the most 3' end of the 113 kD protein was used in a Northern analysis, a single 4.8 KB mRNA species was observed (FIG. 9B).

No independent assay is known for the activity of the 113 kD (or indeed any of the ISGF-3α proteins,) but it is known that the protein is part of a DNA binding complex that can be detected by an electrophoretic mobility shift assay [Fu et al., *PROC. NATL. ACAD. SCI. USA,* 87 (1990)]. Antibodies to DNA binding proteins are known to affect the formation or migration of such complexes. Therefore antiserum to a polypeptide segment (amino acid residues 323 to 527) fused with bacterial glutathione synthetase [Smith et al., *PROC. NATL. ACAD. SCI. USA,* 83 (1986)] was raised in rabbits to determine the reactivity of the ISGF-3 proteins with the antibody. A Western blot analysis showed that the antiserum reacted predominantly with a 113 kD protein both in the ISGF3 fraction purified by specific DNA affinity chromatography (Lane 1) and in crude cell extract (Lane 2, FIG. 10A). The weak reactivity to lower protein bands was

possibly due to 113 kD protein degradation. Most importantly, the antiserum specifically removed almost all of the gel-shift complex leaving some of the oligonucleotide probe in "shifted-shift" complexes which were specifically competed away with a 50 fold molar excess of the oligonucleotide binding site (the ISRE, ref. 2) for ISGF3 (FIG. **10**B). Notably, this antiserum had no effect on the faster migrating shift band produced by ISGF3-γ component alone (FIG. **10**B). Thus it appeared that the antiserum to the 113 kD fusion product does indeed react with another protein that is part of the complete ISGF-3 complex.

A detailed sequence comparison between the 113 and 91 sequences followed (FIG. **8**B): while the nucleotide sequence showed only a distant relationship between the two proteins, there were long stretches of amino acid identity. These conserved regions were scattered throughout almost the entire 715 amino acid length encoded by the 91/84 clone. It was particularly striking that the regions corresponding to amino acids 1 to 48 and 317 to 353 and 654 to 678 in the 113 sequence were 60% to 70% identical to corresponding regions of the 91 kD sequence. Thus the genes encoding the 113 and 84/91 proteins are closely related but not identical.

Through examination for possible consensus sequences that might reveal sub-domain structures in the 113 kD or 84/91 kD sequence, it was found that both proteins contained regions whose sequence might form a coil structure with heptad leucine repeats. This occurred between amino acid 210 and 245 in the 113 kD protein and between 209 and 237 in the 84/91 protein. In both the 113 kD and the 91/84 kD sequences, 4 out of 5 possible heptad repeats were leucine and one was valine. Domains of this type might provide a protein surface that encourages homo-or heterotypic protein interactions which have been observed in several other transcription factors [Vinson et al., *SCIENCE,* 246 (1989)]. An extended acidic domain was located at the carboxyl terminal of the 113 kD protein but not in 91 kD protein (FIG. **8**A), possibly implicating the 113 kD protein in gene activation [Hope et al., Ma et al., *CELL,* 48 (1987)].

Discussion

When compared at moderate or high stringency to the Genbank and EMBL data bases, there were no sequences like 113 or the 84/91 sequence. Preliminary PCR experiments however indicate that there are other family members with different sequences recoverable from a human cell cDNA library (Qureshi and Darnell unpublished). Thus, it appears that the 113 and 84/91 sequences may represent the first two members to be cloned of a larger family of proteins. We would hypothesize that the 113 kD and 84/91 kD proteins may act as signal transducers, somehow interacting with the internal domain of a liganded IFNα receptor or its associated protein and further that a family of waiting cytoplasmic proteins exist whose purpose is to be specific signal transducers when different receptors are occupied. Many experiments lie ahead before this general hypothesis can be crucially tested. Recent experiments have indicated that inhibitors of protein kinases can prevent ISGF-3 complex formulation [Reich et al., *PROC. NATL. ACAD. SCI. USA,* 87 (1990); Kessler et al., *J. BIOL. CHEM.,* 266 (1991)].

However, neither the IFNα or IFNγ receptors that have so far been cloned have intrinsic kinase activity [Uze et al., *CELL,* 60 (1990); Aguet et al., CELL, 55 (1988)]. We would speculate that either a second receptor chain with kinase activity or a separate kinase bound to a liganded receptor could be a part of a complex that would convey signals to the ISGF-3α proteins at the inner surface of the plasma membrane.

From the above, it has been concluded that accurate peptide sequence from ISGF-3 protein components have been determined, leading to correct identification of cDNA clones encoding the 113, 91 and 84 kD components of ISGF-3. Since staurosporine, a broadly effective kinase inhibitor blocks IFN-α induction of transcription and of ISGF-3 formation [Reich et al., *PROC. NATL. ACAD. SCI. USA,* 87 (1990); Kessler et al.,.*J. BIOL. CHEM.,* 266 (1991)] it seems possible that the ISGF-3α proteins are direct cytoplasmic substrates of a liganded receptor-associated kinase. The antiserum against these proteins should prove invaluable in identifying the state of the ISGF-3α proteins before and after IFN treatment and will allow the direct exploration of the biochemistry of signal transduction from the IFN receptor.

EXAMPLE 3

As mentioned earlier, the observation and conclusion underlying the present invention were crystallized from a consideration of the results of certain investigations with particular stimuli. Particularly, the present disclosure is illustrated by the results of work on protein factors that govern transcriptional control of IFNα-stimulated genes, as well as more recent data on the regulation of transcription of genes stimulated by IFNγ.

For example, there is evidence that the 91 kD protein is the tyrosine kinase target when IFNγ is the ligand. Thus two different ligands acting through two different receptors both use these family members. With only a modest number of family members and combinatorial use in response to different ligands, this family of proteins becomes an even more likely possibility to represent a general link between ligand-occupied receptors and transcriptional control of specific genes in the nucleus.

Further study of the 113, 91 and 84 kD proteins of the present invention has revealed that they are phosphorylated in response to treatment of cells with IFNα (FIG. **11**). Moreover, when the phosphoamino acid is determined in the newly phosphorylated protein the amino acid has been found to be tyrosine (FIG. **12**). This phosphorylation has been observed to disappear after several hours, indicating action of a phosphatase of the 113, 91 and 84 kD proteins to stop transcription. These results show that IFN dependent transcription very likely demands this particular phosphorylation and a cycle of interferon-dependent phosphorylation-dephosphorylation is responsible for controlling transcription.

It is proposed that other members of the 113–91 protein family will be identified as phosphorylation targets in response to other ligands. If as is believed, the tyrosine phosphorylation site on proteins in this family is conserved, one can then easily determine which family members are activated (phosphorylated), and likewise the particular extracellular polypeptide ligand to which that family member is responding. The modifications of these proteins (phosphorylation and dephosphorylation) enables the preparation and use of assays for determining the effectiveness of pharmaceuticals in potentiating or preventing intracellular responses to various polypeptides, and such assays are accordingly contemplated within the scope of the present invention.

EXAMPLE 4

Identification of Murine 91 kd Protein

A fragment of the gene encoding the human 91 kD protein was used to screen a murine thymus and spleen cDNA

library for homologous proteins. The screening assay yielded a highly homologous gene encoding a murine polypeptide that is greater than 95% homologous to the human 91 kD protein. The nucleic acid and deduced amino acid sequence of the murine 91 kD protein are shown in FIGS. **13A–13C**, and SEQ ID NO:7 (nucleotide sequence) and SEQ ID NO:8 (amino acid sequence).

EXAMPLE 5

Additional Members of The 113-91 Protein Family

Using a 300 nuclide fragment amplified by PCR from the SH2 region of the murine 91 kD protein gene, murine genes encoding two additional members of the 113–91 family of receptor recognition factor proteins were isolated from a murine splenic/thymic cDNA library according to the method of Sambrook et al. (1989, *Molecular Cloning, A Laboratory Manual,* 2nd. ed., Cold Spring Harbor Press: Cold Spring Harbor, N.Y.) constructed in the ZAP vector. Hybridization was carried out at 42° C. and washed at 42° C. before the first exposure (Church and Gilbert, 1984, *Proc. Natl. Acad. Sci. USA* 81:1991–95). Then the filters were washed in 2×SSC, 0.1% SDS at 65° C. for a second exposure. Stat1 clones survived the 65° C. washing, whereas Stat3 and Stat4 clones were identified as plaques that lost signals at 65° C. The plaques were purified and subcloned according to Stratagene commercial protocols.

This probe was chosen to screen for other STAT family members because, while Stat1 and Stat2 SH2 domains are quite similar over the entire 100 to 120 amino acid region, only the amino terminal half of the STAT SH2 domains strongly resemble the SH2 regions found in other proteins.

The two genes have been cloned into plasmids 13sf1 and 19sf6. The nucleotide sequence, and deduced amino acid sequence, for the 13sf1 and 19sf6 genes are shown in FIGS. **14** and **15**, respectively. These proteins are alternatively termed Stat4 and Stat3, respectively.

Comparison with the sequence of Stat91 (Stat1) and Stat113 (Stat2) shows several highly conserved regions, including the putative SH3 and SH2 domains. The conserved amino acid stretches likely point to conserved domains that enable these proteins to carry out transcription activation functions. Stat3, like Stat1 (Stat91), is widely expressed, while Stat4 expression is limited to the testes, thymus, and spleen. Stat3 has been found to be activated as a DNA binding protein through phosphorylation on: tyrosine in cells treated with EGF or IL-6, but not after IFN-γ, treatment.

Both the 13sf1 and 19sf6 genes share a significant homology with the genes encoding the human and murine 91 kD protein. There is corresponding homology between the deduced amino acid sequences of the 13sf1 and 19sf6 proteins and the amino acid sequences of the human and murine 91 kD proteins, although not the greater than 95% amino acid homology that is found between the murine and human 91 kD proteins. Thus, though clearly of the same family as the 91 kD protein, the 13sf1 and 19sf6 genes encode distinct proteins.

The chromosomal locations of the murine STAT proteins (1–4) have been determined: Stat1 and Stat4 are located in the centromeric region of mouse chromosome 1 (corresponding to human 2q 32–34q); the two other genes are on other chromosomes.

Southern analysis using probes derived from 13sf1 and 19sf6 on human genomic libraries have established that genes corresponding to the murine 13sf1 and 19sf6 genes are found in humans.

Tissue distribution of mRNA expression of these genes was evaluated by Northern hybridization analysis. The results of this distribution analysis are shown in the following Table.

TABLE

DISTRIBUTION OF mRNA EXPRESSION OF 13sf1, 19sf6, 91 kD PROTEINS

ORGAN	13sf1	19sf6	91 KD
BRAIN	–	+	–
HEART	–	+++	–
KIDNEY	–	–	–
LIVER	–	+	+
LUNG	–	–	–
SPLEEN	+	+	++++
TESTIS	++++	++	N.A.
THYMUS	++	++	+++
EMBRYO (16d)	not found	found	found

Northern analysis demonstrates that there is variation in the tissue distribution of expression of the mRNAs encoded by these genes. The variation and tissue distribution indicates that the specific genes encode proteins that are responsive to different factors, as would be expected in accordance with the present invention. The actual ligand, the binding of which induces phosphorylation of the newly discovered factors, will be readily determinable based on the tissue distribution evidence described above.

To determine whether the Stat3 and Stat4 proteins were present in cells, protein blots were carried out with antisera against each protein. The antisera were obtained by subcloning amino acids 688 to 727 of Sta3 and 678 to 743 of Stat4 to pGEX1λ.1 (Pharmacia) by PCR with oligonucleotides based on the boundary sequence plus restriction sites (BamHI at the 5' end and EcoRI at the 3' end), allowing for in-frame fusion with GST. One milligram of each antigen was used for the immunization and three booster injections were given 4 weeks apart. Anti-Stat3 and anti-Stat4 sera were used 1:1000 in Western blots using standard protocols. To avoid cross reactivity of the antisera, antibodies were raised against the C-terminal of Stat3 and Stat4, the less homologous region of the protein.

These proteins were unambiguously found in several tissues where the mRNA wan known to be present. Protein expression was checked in several cell lines as well. A protein of 89 kD reactive with Stat4 antiserum was expressed in 70Z cells, a preB cell line, but not in many other cell lines. Stat3 was highly expressed, predominantly as a 97 kD protein, in 70Z, HT2 (a mouse helper T cell clone), and U937 (a macrophage-derived cell).

To prove that the full length functional cDNA clones of Stat3 and Stat4 were obtained, the open reading frames of each cDNA was independently (i.e., separately) cloned into the Rc/CMV expression vector (Invitrogen) downstream of a CMV promoter. The resulting plasmids were transfected into COS1 cells and proteins were extracted 60 hrs post-transfection and examined by Western blot after electrophoresis. Untransfected COS1 cells expressed a low level of 97 kD Stat3 protein but did not express a detectable level of Stat4. Upon transfection of the Stat3-expressing plasmid, the 97 kD Stat3 was increased at least 10-fold. And 89 kD protein antigenically related to Stat3, found as a minor band in most cell line extracts, was also increased post-transfection. This protein therefore appears to represent another form of Stat3 protein, or an antigenically similar protein whose synthesis is stimulated by Stat3. Transfection with Stat4 led to the expression of a 89 kD reactive band indistinguishable in size form the p89 Stat4 found in 70Z cell extracts.

Discussion

As mentioned earlier, the observation and conclusion underlying the present invention were crystallized from a consideration of the results of certain investigations with particular stimuli. Particularly, the present disclosure is illustrated by the results of work on protein factors that govern transcriptional control of IFNα-stimulated genes, as well as more recent data on the regulation of transcription of genes stimulated by IFN-γ. The present disclosure is further illustrated by the identification of related genes encoding protein factors responsive to as yet unknown factors. It is expected that the murine 91 kD protein is responsive to IFN-γ.

For example, the above represents evidence that the 91 kD protein is the tyrosine kinase target when IFNγ is the ligand. Thus two different ligands acting through two different receptors both use these family members. With only a modest number of family members and combinatorial use in response to different ligands, this family of proteins becomes an even more likely possibility to represent a general link between ligand-occupied receptors and transcriptional control of specific genes in the nucleus.

It is proposed and shown by the foregoing that other members of the 113–91 protein family will be and have been identified as phosphorylation targets in response to other ligands. If as is believed, the tyrosine phosphorylation site on proteins in this family is conserved, one can then easily determine which family members are activated (phosphorylated), and likewise the particular extracellular polypeptide ligand to which that family member is responding. The modifications of these proteins (phosphorylation and dephosphorylation) enables the preparation and use of assays for determining the effectiveness of pharmaceuticals in potentiating or preventing intracellular responses to various polypeptides, and such assays are accordingly contemplated within the scope of the present invention.

Earlier work has concluded that DNA binding protein was activated in the cell cytoplasm in response to IFN-γ treatment and that this protein stimulated transcription of the GBP gene (10,14). In the present work, with the aid of antisera to proteins originally studied in connection with IFN-α gene stimulation (7,12,15), the 91 kD ISGF-3 protein has been assigned a prominent role in IFN-γ gene stimulation as well. The evidence for this conclusion included: 1) antisera specific to the 91 kD protein affected the IFN-γ dependent gel-shift complex, and 2) A 91 kD protein could be cross-linked to the GAS IFN-γ activated site. 3) A ^{35}S-labeled 91 kD protein and a 91 kD immunoreactive protein specifically purified with the gel-shift complex. 4) The 91 kD protein is an IFN-γ dependent tyrosine kinase substrate as indeed it had earlier proved to be in response to IFN-α (15). 5) The 91 kD protein but not the 113 kD protein moved to the nucleus in response to IFN-γ treatment. None of these experiments prove but do strongly suggest that the same 91 kD protein acts differently in different DNA binding complexes that are triggered by either IFN-α or IFN-γ.

These results strongly support the hypothesis originated from studies on IFN-α that polypeptide cell surface receptors report their occupation by extracellular ligand to latent cytoplasmic proteins that after activation move to the nucleus to trigger transcription (4,15,21). Furthermore, because cytoplasmic phosphorylation and factor activation is so rapid it appears likely that the functional receptor complexes contain tyrosine kinase activity. Since the IFN-γ receptor chain that has been cloned thus far (22) has no hint of possessing intrinsic kinase activity, perhaps some other molecule with tyrosine kinase activity couples with the IFN-γ receptor. Two recent results with other receptors suggest possible parallels to the situation with the IFN receptors. The trk protein which has an intracellular tyrosine kinase domain, associates with the NGF receptor when that receptor is occupied (23). In addition, the lck protein, a member of the src family of tyrosine kinases, is co-precipitated with the T cell receptor (24). It is possible to predict that signal transduction to the nucleus through these two receptors could involve latent cytoplasmic substrates that form part of activated transcription factors. In any event, it seems possible that there are kinases like trk or lck associated with the IFN-γ receptor or with IFN-α receptor.

With regard to the effect of phosphorylation on the 91 kD protein, it was something of a surprise that after IFN-γ treatment the 91 kD protein becomes a DNA binding protein. Its role must be different in response to IFN-α treatment. Tyrosine is also phosphorylated on tyrosine and joins a complex with the 113 and 84 kD proteins but as judged by UV cross-linking studies (7), the 91 kD protein does not contact DNA.

In addition to becoming a DNA binding protein it is clear that the 91 kD protein is specifically translocated the nucleus in the wake of IFN-γ stimulation.

EXAMPLE 6

Dimerization of Phosphorylated Stat91

Stat91 (a 91 kD protein that acts as a signal transducer and activator of transcription) is inactive in the cytoplasm of untreated cells but is activated by phosphorylation on tyrosine in response to a number of polypeptide ligands including IFN-α and IFN-γ. This example reports that inactive Stat91 in the cytoplasm of untreated cells is a monomer and upon IFN-γ induced phosphorylation it forms a stable homodimer. The dimer is capable of binding to a specific DNA sequence directing transcription. Dissociation and reassociation assays show that dimerization of Stat91 is mediated through SH2-phosphotyrosyl peptide interactions. Dimerization involving SH2 recognition of specific phosphotyrosyl peptides may well provide a prototype for interactions among family members of STAT proteins to form different transcription complexes and Jak2 for the IFN-γ pathway (42, 43, 44). These kinases themselves become tyrosine phosphorylated to carry out specific signaling events.

Materials and Methods

Cell Culture. Human 2fTGH, U3A cells were maintained in DMEM medium supplied with 10% bovine calf serum. U3A cell lines supplemented with various Stat91 protein constructs were maintained in 0.1 mg/ml G418 (Gibco, BRL).

Stable cell lines were selected as described (45). IFN-γ(5 ng/ml, gift from Amgen) treatment of cells was for 15 min. unless otherwise noted.

Plasmid Constructions. Expression construct MNC-84 was made by insertion of the cDNA into the Not I-Bam HI cloning site of an expression vector PMNC (45, 35). MNC-91L was made by insertion of the Stat91 cDNA into the Not I-Bam HI cloning sites of pMNC without the stop codon at the end, resulting the production of a long form of Stat91 with a C-terminal tag of 34 amino acids encoded by PMNC vector.

GST fusion protein expression plasmids were constructed by the using the pGEX-2T vector (Pharmacia). GST-91SH2

encodes amino acids 573 to 672 of Stat91; GST-91mSH2 encodes amino acids 573 to 672 of Stat91 with an Arg-602—>Leu-602 mutation; and GST-91SH3 encodes amino acids 506 to 564 of Stat91.

DNA Transfection. DNA transfection was carried by the calcium phosphate method, and stable cell lines were selected in Dulbecco's modified Eagle's medium containing G418 (0.5 mg/ml, Gibco), as described (45).

Preparation of Cell Extracts. Crude whole cell extracts were prepared as described (31). Cytoplasmic and nuclear extracts were prepared essentially as described (46).

Affinity Purification. Affinity purification with a biotinylated oligonucleotide was described (31). The sequence of the biotinylated GAS oligonucleotide was from the Ly6E gene promoter (34).

Nondenaturing Polyacrylamide Gel Analysis. A nondenatured protein molecular weight marker kit with a range of molecular weights from 14 to 545 kD was obtained from Sigma. Determining molecular weights using nondenaturing polyacrylamide gel was carried out following the manufacturer's procedure, which is a modification of the methods of Bryan and Davis (47, 48). Phosphorylated and unphosphorylated Stat91 samples obtained from affinity purification using a biotinylated GAS oligonucleotide (31) were resuspended in a buffer containing 10 mM Tris (pH 6.7), 16% glycerol, 0.04% bromphenol blue (BPB). The mixtures were analyzed on 4.5%, 5.5%, 6.5%, and 7.5.% native gels side by side with standard markers using a Bio-Rad mini-Protean II Cell electrophoresis system. Electrophoresis was stopped when the dye (BPB) reached the bottom of the gels. The molecular size markers were revealed by Coomassie blue staining. Phosphorylated and unphosphorylated Stat19 samples were detected by immunoblotting with anti-91T.

Glycerol Gradient Analysis. Cells extracts (Bud 8) were mixed with protein standards (Pharmacia) and subjected to centrifugation through preformed 10%–40% glycerol gradients for 40 hours at 40,000 rpm in an SW41 rotor as described (6).

Gel Mobility Shift Assays. Gel mobility shift assays were carried out as described (34). An oligonucleotide corresponding to the GAS element from the human FcγRI receptor gene (Pearse et al. 1993) was synthesized and used for gel mobility shift assays. The oligonucleotide has the following sequence:

5'GATCGAGATGTATTrCCCAGAAAAG3' (SEQ. ID NO: 17).

Synthesis of Peptides. Solid phase peptide synthesis was used with either a DuPont RAMPS multiple synthesizer or by manual synthesis. C-terminal amino attached to Wang resin were obtained from DuPont/NEN. All amino acids were coupled as the N-Fmoc pentafluorophenyl esters (Advanced Chemtech), except for N-Fmoc, PO-dimethyl-L-phosphotyrosine (Bachem). Double couplings were used. Cleavage from resin and deprotection used thioanisol/m-cresol/TFA/TMSBr at 4° C. for 16 hr. Purification used C-18 column HPLC with 0.1% TFA/acetonitrile gradients. Peptides were characterized by ^{1}H and ^{31}P NMR, and by Mass Spec, and were greater than 95% pure.

Guanidium Hydrochloride Treatment. Extracts were incubated with guanidium hydrochloride (final concentration was 0.4 to 0.6M) for two min. at room temperature and then diluted with gel shift buffer (final concentration of guanidium hydrochloride was 100 mM) and incubated at room temperature for 15 min. ^{32}P-labeled GAS oligonucleotide probe was then added directly to the mixture followed by gel mobility shift assay.

Dissociation-reassociation Analysis. Extracts were incubated with various concentrations of peptides or fusion proteins, and ^{32}P-labeled GAS oligonucleotide probe in gel shift buffer was then added to promote the formation of protein-DNA complex followed by mobility shift analysis. This assay did not involve guanidium hydrochloride treatment.

Preparation of Fusion Proteins. Bacterially expressed GST fusion proteins were purified using standard techniques, as described in Birge et al., 1992. Fusion proteins were quantified by O.D. absorbance at 280 nm. Aliquotes were frozen at −70° C.

Results

Detection of Ligand Induced Dimer Formation of Stat91 in Solution. In untreated cells, Stat91 is not phosphorylated on tyrosine. Treatment with IFN-γ leads within minutes to tyrosine phosphorylation and activation of DNA binding capacity. The phosphorylated form migrates more slowly during electrophoresis under denaturing conditions affording a simple assay for the phosphoprotein (31).

To determine the native molecular weights of the phosphorylated and unphosphorylated forms of Stat91, we separated them by affinity purification using a biotinylated deoxyoligonucleotide containing a GAS sequence (interferon gamma activation site) (FIG. 16A). The separation of phosphorylated Stat91 from the unphosphorylated form was efficient as almost all detectable phosphorylated form could bind to the GAS site while unphosphorylated Stat91 remained unbound. To determine the molecular weights of the purified phosphorylated Stat91 and unphosphorylated Stat91, samples of each were then subjected to electrophoresis through a set of nondenaturing gels containing various concentrations of acrylamide followed by Western blot analysis (FIG. 16B). Native protein size markers (Sigma) were included in the analysis.

This technique was originally described by Bryan (48) and was recently used for dimer analysis (49). The logic of the technique is that increasing gel concentrations affect the migration of larger proteins more than smaller proteins, and the analysis is not affected by modifications such as protein phosphorylation (49).

A function of the relative mobilities (Rm) was plotted versus the concentration of acrylamide for each sample to construct Ferguson plots (FIG. 16C). The logarithm of the retardation coefficient (calculated from FIG. 16C) of each sample was then plotted against the logarithm of the relevant molecular weight range (FIG. 16D). By extrapolation of its retardation coefficient (FIG. 16D), the native molecular weight of Stat91 from untreated cells was estimated to be approximately 95 kD, while tyrosine phosphorylated Stat91 was estimated to be about twice as large, or approximately 180 kD. Because the calculated molecular weight from amino acid sequence of Stat91 is 87 kD, and Stat91 migrates on denaturing SDA gels with an apparent molecular weight of 91 kD (see supra, and refs. 12 and 45), we concluded that in solution, unphosphorylated Stat91 existed as a monomer while tyrosine phosphorylated Stat91 is a dimer.

We also employed glycerol gradient analysis to estimate the native molecular weights of both phosphorylated and unphosphorylated Stat91 (FIG. 17). Whole cell extract of fibroblast cells (Bud8) treated with IFN-γ were prepared and subjected to sedimentation through a 10–40% glycerol gradient. Fractions from the gradient were collected and analyzed by both immunoblotting and gel mobility shift analysis (FIGS. 17A and 17B). As expected, two electrophoretic

forms of Stat91 could be detected by immunoblotting (FIG. 17A): the slow-migrating form (tyrosine phosphorylated) and the fast-migrating form (unphosphorylated; FIG. 17A). The phosphorylated Stat91 sedimented more rapidly than the unphosphorylated form. Again, using molecular weight markers, the native molecular weight of the unphosphorylated form of Stat91 appeared to be about 90 kD while the tyrosine phosphorylated form of Stat91 was about 180 kD (FIG. 17C), supporting the conclusion that unphosphorylated Stat91 existed as a monomer in solution while the tyrosine phosphorylated form exists as a dimer. When fractions from the glycerol gradients were analyzed by electrophoretic mobility shift analysis (FIG. 17B), the peak of the phosphorylated form of Stat91 correlated well with the DNA-binding activity of Stat91. Thus only the phosphorylated dimeric Stat91 has the sequence-specific DNA recognition capacity.

Stat91 Binds DNA as a Dimer. Long or short versions of DNA binding protein can produce, respectively, a slower or a faster migrating band during gel retardation assays. Finding intermediate gel shift bands produced by mixing two different sized species provides evidence of dimerization of the DNA binding proteins. Since Stat91 requires specific tyrosine phosphorylation in ligand-treated cells for its DNA binding, we sought evidence of formation of such heterodimers, first in transfected cells. An expression vector (MNC911) encoding Stat91L, a recombinant form of Stat91 containing an additional 34 amino acid carboxyl terminal tag was generated. [The extra amino acids were encoded by a segment of DNA sequence from plasmid pMNC (see Materials and Methods).] A Stat84 expression vector (MNC84) was also available (45). From somatic cell genetic experiments, mutant human cell lines (U3) are known that lack the Stat91/84 mRNA and proteins (29,30). The U3 cells were therefore separately transfected with vectors encoding Stat84 (MNC84) or Stat91L (MNC91L) or a mixture of both vectors. Permanent transfectants expressing Stat84 (C84), Stat91L (C91L) or both proteins (Cmx) were isolated (FIG. 18A).

Mobility shift analysis was performed with extracts from these stable cell lines (FIG. 18B). Extracts of IFN-γ-treated C84 cells produced a faster migrating gel shift band than extracts of treated C91L cells. Most importantly, extracts from IFN-γ-treated Cmx cells expressing both Stat84 and Stat91L proteins formed an additional intermediate gel shift band. Anti-91, an antiserum against the C-terminal 38 amino acids of Stat91 (12) that are absent in Stat84, specifically removed the top two shift bands seen with the Cmx extracts. Anti-91, an antiserum against amino acids 609 to 716 (15) that recognizes both Stat91L and Stat84, proteins inhibited the binding of all three shift bands. Thus, the middle band formed by extracts of the Cmx cells is clearly identified as a heterodimer of Stat84 and Stat91L. We concluded that both Stat91 and Stat84 bind DNA as homodimers and, if present in the same cell, win form heterodimers.

We next wanted to detect the formation of dimers in vitro. When cytoplasmic or nuclear extracts of IFN-γ-treated C84 or C91L cells were mixed and analyzed (FIG. 19), only the fast or slow migrating gel shift bands were observed. Thus it appeared that once formed in vivo, the dimers were stable. To promote the formation of protein interchange between the subunits of the dimer, a mixture of either cytoplasmic or nuclear extracts of IFN-γ-treated C84 or C91L cells were subjected mild denaturation-renaturation treatment: extracts were made 0.5M with respect to guanidium hydrochloride for two minutes and then diluted for renaturation and subsequently used for gel retardation analysis. The formation of heterodimer was clearly detected after this treatment. When extracts from either C84 cells alone or C91L cells alone were subjected to the same treatment, the intermediate band did not form. The intermediate band was again proven by antiserum treatment to consist of Stat84/Stat91L dimer (data not shown).

This experiment defined conditions under which the dimer was stable, but also showed that dissociation and reassociation of the dimer in vitro was possible. Since guanidium hydrochloride is known to disrupt only non-covalent chemical bonds, it seemed that Stat91 (or Stat84) homodimerization was mediated through non-covalent interactions.

Dimerization of Stat91 Involves Phosphotyrosyl Peptide and SH2 Interactions. Based on the results described above, we devised a dissociation-reassociation assay in the absence of guanidium hydrochloride to explore the possible nature of interactions involved in dimer formation (FIG. 20). When the short and the long forms of a homodimer are mixed with a dissociating agent (e.g., a peptide containing the putative dimerization domain), the subunits of the dimer should dissociate (in a concentration dependent fashion) due to the interaction of the agent with the dimerization domain(s) of the protein. When a specific DNA probe is subsequently added to the mixture to drive the formation of a stable protein-DNA complex, the detection of any reassociated or remaining dimers can be assayed. In the presence of low concentration of the dissociating agent, addition of DNA to form the stable protein-DNA complex should lead to the detection of homodimers as well as heterodimers. At high concentration of the dissociating agent, subunits of the dimer may not be able to re-form and no DNA-protein complexes would be detected (FIG. 20).

The Stat91 sequence contains an SH2 domain (amino acids 569 to 700, see discussion below), and we knew that Tyr-701 was the single phosphorylated tyrosine residue required for DNA binding activity (supra, 45). Furthermore, we have observed that phosphotyrosine at 10 mM, but not phosphoserine or phosphothreonine, could prevent the formation of Stat91-DNA complex. We therefore sought evidence that the dimerization of Stat91 involved specific SH2-phosphotyrosine interaction using the dissociation and reassociation assay.

In order to evaluate the role of the SH2-phosphotyrosine interation, two peptides fragments of Stat91 corresponding to segments of the SH2 and phosphotyrosing domains of Stat91 were prepared: a non-phosphorylated peptide (91Y), LDGPKGTGYIKTELI (SEQ. ID NO: 18) (corresponding to amino acids 693–707), and a phosphotyrosyl peptide (91Y-p), GY*IKTE (SEQ. ID NO: 19) (representing residues 700–705).

Activated Stat84 or Stat91L was obtained from IFN-γ-treated C84 or C91L cells and mixed in the presence of various concentrations of the peptides followed by gel mobility shift analysis. The non-phosphorylated peptide had no effect on the presence of the two gel shift bands characteristic of Stat84 or Stat91L homodimers (FIG. 21, lane 2–4). In contrast, the phosphorylated peptide (91Y-p) at the concentration of 4 μM clearly promoted the exchange between the subunits of Stat84 dimers and Stat91L dimers to form heterodimers (FIG. 21, lane 5). At a higher concentration (160 μM), peptide 91Y-p but not the unphosphorylated peptide dissociated the dimers and blocked the formation of DNA protein complexes (FIG. 21, lane 7).

When cells are treated with IFN-α both Stat91 (or 84) and Stat113 become phosphorylated (15). Antiserum to Stat113

can precipitate both Stat113 and Stat91 after IFN-α-treatment but not before, suggesting IFN-α dependent interaction of these two proteins, perhaps as a heterodimer (15).

In Stat113, tyr-690 in the homologous position to Tyr-701 in Stat91 is the single target residue for phosphorylation. Amino acids downstream of the affected tyrosine residue show some homology between the two proteins. We therefore prepared a phosphotyrosyl peptide of Stat113 (113Y-p), KVNLQERRKY*LKHR (SEQ. ID NO:20) [amino acids 681 to 694; (38)]. At concentrations similar to 91Y-p, 113Y-p also promoted the exchange of subunits between the Stat84 and Stat91L, while at a high concentration (40 μM), 113Y-p prevented the gel shift bands almost completely (FIG. **21**, lane 8–10).

We prepared a phosphotyrosyl peptide (SrcY-p), EPQY*EEIPIYL (SEQ. ID NO:21) which is known to interact with the Src SH2 domain with a high affinity (50). This peptide showed no effect on the Stat91 dimer formation (FIG. **21**, lane 11–13). Thus, it seems that Stat91 dimerization involves SH2 interaction with tyrosine residues in specific peptide sequence.

To test further the specificity of Stat91 dimerization mediated through specific-phosphotyrosyl-peptide SH2 interaction, a fusion product of glutathione-S-transferase with the Stat91-SH2 domain (GST-91SH2) was prepared (FIG. **22**A) and used in the in vitro dissociation reassociation assay. At concentrations of 0.5 to 5 μM, the Stat91-SH2 domain promoted the formation of a heterodimer (FIG. **22**B, lanes 5–7). In contrast, neither GST alone, nor fusion products with a mutant (R^{602}—→L^{602}) Stat91-SH2 domain (GST 91mSH2) that renders Stat91 non-functional in vivo, a Stat91 SH3 domain (GST-91SH3), nor the Src SH2 domain (GST-SrcSH2), induced the exchange of subunits between the Stat84 and Stat91L homodimers (FIG. **22**B).

Discussion

The initial sequence analysis of the Stat91 and Stat113 proteins revealed the presence of SH2 like domains (see 13,38). Further it was found that STAT proteins themselves are phosphorylated on single tyrosine residues during their activation (15,31). Single amino acid mutations either removing the Stat91 phosphorylation site, Tyr-701, or converting Arg-702 to Leu in the highly conserved "pocket" region of the SH2 domain abolished the activity of Stat91 (45). Thus it seemed highly likely that one possible role of the STAT SH2 domains would be to bind the phosphotyrosine residues in one of the JAK kinases.

Since the activated STATs have phosphotyrosine residues and SH2 domains, a second suggested role for SH2 domains was in protein-protein interactions within the STAT family. By two physical criteria—electrophoresis in native gels and sedimentation on gradients—Stat91 in untreated cells is a monomer and in treated cells is a dimer (FIGS. **16–18**). Since phosphotyrosyl peptides from Stat91 or Stat113 and the SH2 domain of Stat91 could efficiently promote the formation of herterodimers between Stat91L and Stat84 in a disassociation and reassociation assay, we conclude that dimerization of Stat91 involves SH2-phosphotyrosyl peptide interactions.

The possibility of an SH2 domain in Stat91 was indicated initially by the presence of highly conserved amino acid stretches between the Stat91 and Stat113 sequences in the 569 to 700 residue region, several of which, especially the FLLR sequence in the amino terminal end of the region, are characteristic of –SH2 domains. The C-terminal half of the SH2 domains are less well conserved in general (39); this was also true for the STAT proteins compared to other proteins, although Stat91 and Stat113 are quite similar in this region (38, 13, FIG. **23**). The available structures of lck, src, abl, and p85a SH2's permit identification of structurally conserved regions (SCR's), and detailed alignment of amino acid sequences of several proteins (FIG. **23**) is based on these.

The characteristic W (in βA1) is preceded by hydrophilic residues and is followed by hydrophobic residues in Stat91, but alignment to the W seems justified, even if the small beta sheet of which the W is part is shifted in Stat91. The three positively charged residues contributing to the phosphotyrosyl binding site are at the positions indicated as alphaA2, betaB5, and betaD5. FIG. **23** shows an alignment which accomplishes this by insertions in the 'AA' and 'CD' regions. This is a different alignment from that previously suggested (38), and gives a satisfactory alignment in the (beta)D region, although, like the previous alignment, it is obviously considerably less similar to the other SH2's in the C-terminus.

This alignment suggests that the SH2 domain in the Stat91 would end in the vicinity of residue 700. In such an alignment, the Tyr-701 occurs almost immediately after the SH2 domain: a distance too short to allow an intramolecular phosphotyrosine –SH2 interaction. Since the data presented earlier strongly implicate that an SH2-phosphotyrosine interaction is involved in dimerization, such an interaction is likely to be between two phospho Stat91 subunits as a reciprocal pTyr –SH2 interaction.

The apparent stability of Stat91 dimer may be due to a high association rate coupled with a high dissociation rate of SH2-phosphotyrosyl peptide interactions as suggested (Felder et al., 1993, Mol. Cell Biol. 13:1449–1455) coupled with interactions between other domains of Stat91 that may contribute stability to the Stat91 dimer. Interference by homologous phosphopeptides with the –SH2-phosphotyrosine interaction would then lower stability sufficiently to allow complete dissociation and heterodimerization.

The dimer formation between phospho Stat91 is the first case in eukaryotes where dimer formation is regulated by phosphorylation, and the only one thus far dependent on tyrosine phosphorylation. We anticipate that dimerization with the STAT protein family will be important. It seems likely that in cells treated with IFN-α, there is Stat113-Stat91 interaction (15). This may well be mediated through SH2 and phosphotyrosyl peptide interactions as described above, leading to a complex (a probable dimer of Stat91-Stat113) which joins with a 48 kD DNA binding protein (a member of another family of DNA binding factors) to make a complex capable of binding to a different DNA site. Furthermore, we have recently cloned two mouse cDNAs which encode other STAT family members that have conserved the same general structure features observed in the Stat91 and Stat113 molecules (see Example 5, Supra). (U.S. application Ser. No. 08/126,588, filed Sep. 29, 1993, which is specifically incorporated herein by reference in its entirety). Thus the specificity of STAT-containing complexes will almost surely be affected by which proteins are phosphorylated and then available for dimer formation.

The following is a list of references related to the above disclosure and particularly to the experimental procedures and discussions. The references are numbered to correspond to like number references that appear hereinabove.

1. Lamer, A. C., Jonak, G., Cheng, Y. S., Korant, B., Knight, E. and Darnell, J. E., Jr. (1984). *Proc. Natl. Acad. Sci.*

USA 81:6733–6737; Larner, A. C., Chaudhuri, A. and Darnell, J. E. (1986). *J. Biol. Chem.* 261:453–459.
2. Friedman, R. L., Manly, S. P., McMahon, M., Kerr, I. M. and Stark, G. R. (1984). *Cell* 38:745–755.
3. Levy, D. E., Kessler, D. S., Pine, R., Reich, N. and Darnell, J. E. (1988). *Genes & Dev.* 2:383–392.
4. Levy, D. E., Kessler, D. S., Pine, R., and Darnell, J. E. (1989). *Genes & Dev.* 3:1362–1371.
5. Dale, T. C., Iman, A. M. A., Kerr, I. M. and Stark, G. R. (1989). *Proc. Natl. Acad. Sci.* 86:1203–1207.
6. Kessler, D. S., Veals, S. A., Fu, X. -Y., and Levy, D. E. (1990). *Genes & Dev.* 4:1753–1765.
7. Fu, X. -Y., Kessler, D. S., Veals, S. A., Levy, D. E. and Darnell, J. E. (1990). *Proc. Natl. Acad. Sci. USA* 87:8555–8559.
8. Lew, D. J., Decker, T., and Darnell, J. E. (1989). *Mol. Cell. Biol.* 9:5404–5411.
9. Decker, T., Lew, D. J., Cheng, Y. -S., Levy, D. E. and Darnell, J. E. (1989). *EMBO J.* 8:2009–2014.
10. Decker, T., Lew, D. J., Mirkovitch, J. and Darnell, J. E., 1991. *EMBO J.* 10:927–932.
11. Veals, S. A., Schindler, C. W., Fu, X. -Y., Leonard, D., Darnell, J. E. and Levy, D. E. (1992). *Mol. Cell. Biol.* 12.
12. Schindler, C., Fu, X. -Y., Improta, T., Aebersold, R. and Darnell, J. E. (1992). *Proc. Natl. Acad. Sci. USA* 89:7836–7389.
13. Fu, X. -Y., Schindler, C., Improta, T., Aebersold, R. and Darnell, J. E. (1992). *Proc. Natl. Acad. Sci. USA* 89:7840–7389.
14. Lew, D. J., Decker, T. and Darnell, J. E. (1991). *Mol. Cell. Biol.* 11: 182–191.
15. Schindler, C., Shuai, K., Fu, X. -Y., Prezioso, V. and Darnell, J. E. (1992). *Science* 257:809–812.
16. Garner, M. M. and Revan, A. (1981). *Nuc. Acids Res.* 9:3047–3059; Fried, A., and Crothers, D. M. (1981) ibid 6505–6525.
17. Celis, J. E., Justessen, J., Madsun, P. S., Lovmand, J., Ratz, G. P. and Celis, A. (1987). *Leukemia* 1:800–813.
18. Chodosh, L. A., Carthew, R. W. and Sharp, P. A. (1986). *Mol. Cell Biol.* 6:4723–4733.
19. Reich, N. and Pfeffer, L. M. (1990). *Proc. Natl. Acad. Sci. USA* 87:8761–8765.
20. Kessler, D. S. and Levy, D. (1991). *J. Biol. Chem.* 266: 23471–23476.
21. Levy, D., and Darnell, J. E. (1990). *The New Biologist* 2:923–928.
22. Aguet, J. M., Denbie, Z. and Merlin, G. (1986). *Cell* 55:273–280.
23. Kaplan, D. R., Martin-Zanca, D. and Parada, L. F. (1991). *Nature* 350:158–160; Hempstead, G., Kapland, D., Martin-Zanca, D., Parada, L. F. and Chao, M. (1991). *Nature* 350:678–683.
24. Veillette, A., Bookman, M. A., Horak, E. M., and Bolen, J. B. (1988). *Cell* 55:301–308; Rudd, C. E. et al. (1988). *Proc. Natl. Acad. Sci. USA* 85:5190–5194.
25. Evans, R. K., Johnson, J. D. and Haley, B. E. (1986). *Proc. Natl. Acad. Sci. USA* 83:5382–5386.
26. Walaas, S. I. and Nairn, A. C. (1989). *J. of Mol. Neurosci.* 1:117–127.
28. Pellegrini, S., John, J., Shearer, M., Kerr, I. M., and Stark, G. R. (1989). *Mol. Cell. Biol.* 9:4605.
29. McKendry, R. et al. (1991). *Proc. Natl. Acad. Sci. U.S.A.* 88:11455.
30. Muller, M. et al. (1993), *EMBO. J.*
31. Shuai, K., Schindler, C., Prezioso, V., and Darnell, J. E., Jr. (1992). *Science* 258:1808.
32. Boyle, W. J., van der Geer, and Hunter, T. (1991). *Methods Enzymol.* 202:110.
33. Walaas, S. I., and Nairn, A. C. (1989). *J. Mol. Neurosci.* 1:117.
34. Khan, K. D. et al. (1993). *Proc. Natl. Acad. Sci. U.S.A.* 90:6806.
35. Qufeshi, S. A. et al. (1991). *J. Biol. Chem.* 266:20594.
36. Kawasaki, E., (1990). In *PRC Protcols: A Guide to Methods and Applications.* Innis, M., Gelfand, D., Sinisky, J., & White, T., Eds. (San Diego, Calif.: Academic Press), p. 119.
37. Brasier, A. R., Tata, J. E., and Habener, J. F. (1989). *Biotechniques* 7:1116.
38. Fu, X. -Y. (1992). *Cell* 70:323–335.
39. Koch, C. A., Anderson, D. Moran, M. F., Ellis, C. and Pawson, T. (1991). *Science* 252:668–674.
40. Waksman, G. et al. (1992). *Nature* 358:646–653.
41. Overduin, M. Rios, C. B., Mayer, B. J., Baltimore, D. and Cowburn, D. (1992). *Cell* 70:697–704.
42. Velazquez et al. (1992). *Cell* 70:313.
43. Muller et al. (1993). *Nature* 366:129–135.
44. Watling et al. (1993). *Nature* 366:166.
45. Shuai, K., Stark, G. R., Kerr, I. M., and Darnell, J. E. (1993). *Science* 261:1744.
46. Dignam et al. (1983). *Nucl. Acids Res.* 11:1475.
47. Davis, B. J. (1964). *Ann. N.Y. Acad. Sci.* 121:404.
48. Bryan, J. K. (1977). *Anal. Biochem.* 78:513.
49. Amster-Choder, O. and Wright, A. (1992). *Science* 257:1395.
50. Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, et al. (1993). *Cell* 72:767.

This invention may be embodied in other forms or carried out in other ways without departing from the spirit or essential characteristics thereof. The present disclosure is therefore to be considered as in all respects illustrative and not restrictive, the scope of the invention being indicated by the appended Claims, and all changes which come within the meaning and range of equivalency are intended to be embraced therein.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 25

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3268 base pairs
(B) TYPE: nucleic acid

-continued

(C) STRANDEDNESS: both
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: HeLa

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 25..2577

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ACTGCAACCC TAATCAGAGC CCAA ATG GCG CAG TGG GAA ATG CTG CAG AAT      51
                           Met Ala Gln Trp Glu Met Leu Gln Asn
                             1               5

CTT GAC AGC CCC TTT CAG GAT CAG CTG CAC CAG CTT TAC TCG CAC AGC      99
Leu Asp Ser Pro Phe Gln Asp Gln Leu His Gln Leu Tyr Ser His Ser
 10                  15                  20                  25

CTC CTG CCT GTG GAC ATT CGA CAG TAC TTG GCT GTC TGG ATT GAA GAC     147
Leu Leu Pro Val Asp Ile Arg Gln Tyr Leu Ala Val Trp Ile Glu Asp
                 30                  35                  40

CAG AAC TGG CAG GAA GCT GCA CTT GGG AGT GAT GAT TCC AAG GCT ACC     195
Gln Asn Trp Gln Glu Ala Ala Leu Gly Ser Asp Asp Ser Lys Ala Thr
             45                  50                  55

ATG CTA TTC TTC CAC TTC TTG GAT CAG CTG AAC TAT GAG TGT GGC CGT     243
Met Leu Phe Phe His Phe Leu Asp Gln Leu Asn Tyr Glu Cys Gly Arg
         60                  65                  70

TGC AGC CAG GAC CCA GAG TCC TTG TTG CTG CAG CAC AAT TTG CGG AAA     291
Cys Ser Gln Asp Pro Glu Ser Leu Leu Leu Gln His Asn Leu Arg Lys
     75                  80                  85

TTC TGC CGG GAC ATT CAG CCC TTT TCC CAG GAT CCT ACC CAG TTG GCT     339
Phe Cys Arg Asp Ile Gln Pro Phe Ser Gln Asp Pro Thr Gln Leu Ala
 90                  95                 100                 105

GAG ATG ATC TTT AAC CTC CTT CTG GAA GAA AAA AGA ATT TTG ATC CAG     387
Glu Met Ile Phe Asn Leu Leu Leu Glu Glu Lys Arg Ile Leu Ile Gln
                110                 115                 120

GCT CAG AGG GCC CAA TTG GAA CAA GGA GAG CCA GTT CTC GAA ACA CCT     435
Ala Gln Arg Ala Gln Leu Glu Gln Gly Glu Pro Val Leu Glu Thr Pro
            125                 130                 135

GTG GAG AGC CAG CAA CAT GAG ATT GAA TCC CGG ATC CTG GAT TTA AGG     483
Val Glu Ser Gln Gln His Glu Ile Glu Ser Arg Ile Leu Asp Leu Arg
        140                 145                 150

GCT ATG ATG GAG AAG CTG GTA AAA TCC ATC AGC CAA CTG AAA GAC CAG     531
Ala Met Met Glu Lys Leu Val Lys Ser Ile Ser Gln Leu Lys Asp Gln
    155                 160                 165

CAG GAT GTC TTC TGC TTC CGA TAT AAG ATC CAG GCC AAA GGG AAG ACA     579
Gln Asp Val Phe Cys Phe Arg Tyr Lys Ile Gln Ala Lys Gly Lys Thr
170                 175                 180                 185

 CCC TCT CTG GAC CCC CAT CAG ACC AAA GAG CAG AAG ATT CTG CAG GAA     627
Pro Ser Leu Asp Pro His Gln Thr Lys Glu Gln Lys Ile Leu Gln Glu
                190                 195                 200

ACT CTC AAT GAA CTG GAC AAA AGG AGA AAG GAG GTG CTG GAT GCC TCC     675
Thr Leu Asn Glu Leu Asp Lys Arg Arg Lys Glu Val Leu Asp Ala Ser
            205                 210                 215

AAA GCA CTG CTA GGC CGA TTA ACT ACC CTA ATC GAG CTA CTG CTG CCA     723
Lys Ala Leu Leu Gly Arg Leu Thr Thr Leu Ile Glu Leu Leu Leu Pro
```

-continued

```
        220                 225                 230

AAG TTG GAG GAG TGG AAG GCC CAG CAG CAA AAA GCC TGC ATC AGA GCT      771
Lys Leu Glu Glu Trp Lys Ala Gln Gln Gln Lys Ala Cys Ile Arg Ala
    235                 240                 245

CCC ATT GAC CAC GGG TTG GAA CAG CTG GAG ACA TGG TTC ACA GCT GGA      819
Pro Ile Asp His Gly Leu Glu Gln Leu Glu Thr Trp Phe Thr Ala Gly
250                 255                 260                 265

GCA AAG CTG TTG TTT CAC CTG AGG CAG CTG CTG AAG GAG CTG AAG GGA      867
Ala Lys Leu Leu Phe His Leu Arg Gln Leu Leu Lys Glu Leu Lys Gly
                270                 275                 280

CTG AGT TGC CTG GTT AGC TAT CAG GAT GAC CCT CTG ACC AAA GGG GTG      915
Leu Ser Cys Leu Val Ser Tyr Gln Asp Asp Pro Leu Thr Lys Gly Val
            285                 290                 295

GAC CTA CGC AAC GCC CAG GTC ACA GAG TTG CTA CAG CGT CTG CTC CAC      963
Asp Leu Arg Asn Ala Gln Val Thr Glu Leu Leu Gln Arg Leu Leu His
        300                 305                 310

AGA GCC TTT GTG GTA GAA ACC CAG CCC TGC ATG CCC CAA ACT CCC CAT     1011
Arg Ala Phe Val Val Glu Thr Gln Pro Cys Met Pro Gln Thr Pro His
    315                 320                 325

CGA CCC CTC ATC CTC AAG ACT GGC AGC AAG TTC ACC GTC CGA ACA AGG     1059
Arg Pro Leu Ile Leu Lys Thr Gly Ser Lys Phe Thr Val Arg Thr Arg
330                 335                 340                 345

CTG CTG GTG AGA CTC CAG GAA GGC AAT GAG TCA CTG ACT GTG GAA GTC     1107
Leu Leu Val Arg Leu Gln Glu Gly Asn Glu Ser Leu Thr Val Glu Val
                350                 355                 360

TCC ATT GAC AGG AAT CCT CCT CAA TTA CAA GGC TTC CGG AAG TTC AAC     1155
Ser Ile Asp Arg Asn Pro Pro Gln Leu Gln Gly Phe Arg Lys Phe Asn
            365                 370                 375

ATT CTG ACT TCA AAC CAG AAA ACT TTG ACC CCC GAG AAG GGG CAG AGT     1203
Ile Leu Thr Ser Asn Gln Lys Thr Leu Thr Pro Glu Lys Gly Gln Ser
        380                 385                 390

CAG GGT TTG ATT TGG GAC TTT GGT TAC CTG ACT CTG GTG GAG CAA CGT     1251
Gln Gly Leu Ile Trp Asp Phe Gly Tyr Leu Thr Leu Val Glu Gln Arg
    395                 400                 405

TCA GGT GGT TCA GGA AAG GGC AGC AAT AAG GGG CCA CTA GGT GTG ACA     1299
Ser Gly Gly Ser Gly Lys Gly Ser Asn Lys Gly Pro Leu Gly Val Thr
410                 415                 420                 425

GAG GAA CTG CAC ATC ATC AGC TTC ACG GTC AAA TAT ACC TAC CAG GGT     1347
Glu Glu Leu His Ile Ile Ser Phe Thr Val Lys Tyr Thr Tyr Gln Gly
                430                 435                 440

CTG AAG CAG GAG CTG AAA ACG GAC ACC CTC CCT GTG GTG ATT ATT TCC     1395
Leu Lys Gln Glu Leu Lys Thr Asp Thr Leu Pro Val Val Ile Ile Ser
            445                 450                 455

 AAC ATG AAC CAG CTC TCA ATT GCC TGG GCT TCA GTT CTC TGG TTC AAT     1443
Asn Met Asn Gln Leu Ser Ile Ala Trp Ala Ser Val Leu Trp Phe Asn
        460                 465                 470

TTG CTC AGC CCA AAC CTT CAG AAC CAG CAG TTC TTC TCC AAC CCC CCC     1491
Leu Leu Ser Pro Asn Leu Gln Asn Gln Gln Phe Phe Ser Asn Pro Pro
    475                 480                 485

AAG GCC CCC TGG AGC TTG CTG GGC CCT GCT CTC AGT TGG CAG TTC TCC     1539
Lys Ala Pro Trp Ser Leu Leu Gly Pro Ala Leu Ser Trp Gln Phe Ser
490                 495                 500                 505

TCC TAT GTT GGC CGA GGC CTC AAC TCA GAC CAG CTG AGC ATG CTG AGA     1587
Ser Tyr Val Gly Arg Gly Leu Asn Ser Asp Gln Leu Ser Met Leu Arg
                510                 515                 520

AAC AAG CTG TTC GGG CAG AAC TGT AGG ACT GAG GAT CCA TTA TTG TCC     1635
Asn Lys Leu Phe Gly Gln Asn Cys Arg Thr Glu Asp Pro Leu Leu Ser
            525                 530                 535

TGG GCT GAC TTC ACT AAG CGA GAG AGC CCT CCT GGC AAG TTA CCA TTC     1683
```

-continued

```
Trp Ala Asp Phe Thr Lys Arg Glu Ser Pro Pro Gly Lys Leu Pro Phe
        540                 545                 550

TGG ACA TGG CTG GAC AAA ATT CTG GAG TTG GTA CAT GAC CAC CTG AAG     1731
Trp Thr Trp Leu Asp Lys Ile Leu Glu Leu Val His Asp His Leu Lys
    555                 560                 565

GAT CTC TGG AAT GAT GGA CGC ATC ATG GGC TTT GTG AGT CGG AGC CAG     1779
Asp Leu Trp Asn Asp Gly Arg Ile Met Gly Phe Val Ser Arg Ser Gln
570                 575                 580                 585

GAG CGC CGG CTG CTG AAG AAG ACC ATG TCT GGC ACC TTT CTA CTG CGC     1827
Glu Arg Arg Leu Leu Lys Lys Thr Met Ser Gly Thr Phe Leu Leu Arg
                590                 595                 600

TTC AGT GAA TCG TCA GAA GGG GGC ATT ACC TGC TCC TGG GTG GAG CAC     1875
Phe Ser Glu Ser Ser Glu Gly Gly Ile Thr Cys Ser Trp Val Glu His
            605                 610                 615

CAG GAT GAT GAC AAG GTG CTC ATC TAC TCT GTG CAA CCG TAC ACG AAG     1923
Gln Asp Asp Asp Lys Val Leu Ile Tyr Ser Val Gln Pro Tyr Thr Lys
        620                 625                 630

GAG GTG CTG CAG TCA CTC CCG CTG ACT GAA ATC ATC CGC CAT TAC CAG     1971
Glu Val Leu Gln Ser Leu Pro Leu Thr Glu Ile Ile Arg His Tyr Gln
    635                 640                 645

TTG CTC ACT GAG GAG AAT ATA CCT GAA AAC CCA CTG CGC TTC CTC TAT     2019
Leu Leu Thr Glu Glu Asn Ile Pro Glu Asn Pro Leu Arg Phe Leu Tyr
650                 655                 660                 665

CCC CGA ATC CCC CGG GAT GAA GCT TTT GGG TGC TAC TAC CAG GAG AAA     2067
Pro Arg Ile Pro Arg Asp Glu Ala Phe Gly Cys Tyr Tyr Gln Glu Lys
                670                 675                 680

GTT AAT CTC CAG GAA CGG AGG AAA TAC CTG AAA CAC AGG CTC ATT GTG     2115
Val Asn Leu Gln Glu Arg Arg Lys Tyr Leu Lys His Arg Leu Ile Val
            685                 690                 695

GTC TCT AAT AGA CAG GTG GAT GAA CTG CAA CAA CCG CTG GAG CTT AAG     2163
Val Ser Asn Arg Gln Val Asp Glu Leu Gln Gln Pro Leu Glu Leu Lys
        700                 705                 710

CCA GAG CCA GAG CTG GAG TCA TTA GAG CTG GAA CTA GGG CTG GTG CCA     2211
Pro Glu Pro Glu Leu Glu Ser Leu Glu Leu Glu Leu Gly Leu Val Pro
    715                 720                 725

 GAG CCA GAG CTC AGC CTG GAC TTA GAG CCA CTG CTG AAG GCA GGG CTG    2259
Glu Pro Glu Leu Ser Leu Asp Leu Glu Pro Leu Leu Lys Ala Gly Leu
730                 735                 740                 745

GAT CTG GGG CCA GAG CTA GAG TCT GTG CTG GAG TCC ACT CTG GAG CCT     2307
Asp Leu Gly Pro Glu Leu Glu Ser Val Leu Glu Ser Thr Leu Glu Pro
                750                 755                 760

GTG ATA GAG CCC ACA CTA TGC ATG GTA TCA CAA ACA GTG CCA GAG CCA     2355
Val Ile Glu Pro Thr Leu Cys Met Val Ser Gln Thr Val Pro Glu Pro
            765                 770                 775

GAC CAA GGA CCT GTA TCA CAG CCA GTG CCA GAG CCA GAT TTG CCC TGT     2403
Asp Gln Gly Pro Val Ser Gln Pro Val Pro Glu Pro Asp Leu Pro Cys
        780                 785                 790

GAT CTG AGA CAT TTG AAC ACT GAG CCA ATG GAA ATC TTC AGA AAC TGT     2451
Asp Leu Arg His Leu Asn Thr Glu Pro Met Glu Ile Phe Arg Asn Cys
    795                 800                 805

GTA AAG ATT GAA GAA ATC ATG CCG AAT GGT GAC CCA CTG TTG GCT GGC     2499
Val Lys Ile Glu Glu Ile Met Pro Asn Gly Asp Pro Leu Leu Ala Gly
810                 815                 820                 825

CAG AAC ACC GTG GAT GAG GTT TAC GTC TCC CGC CCC AGC CAC TTC TAC     2547
Gln Asn Thr Val Asp Glu Val Tyr Val Ser Arg Pro Ser His Phe Tyr
                830                 835                 840

ACT GAT GGA CCC TTG ATG CCT TCT GAC TTC TAGGAACCAC ATTTCCTCTG       2597
Thr Asp Gly Pro Leu Met Pro Ser Asp Phe
            845                 850
```

-continued

```
TTCTTTTCAT ATCTCTTTGC CCTTCCTACT CCTCATAGCA TGATATTGTT CTCCAAGGAT   2657

GGGAATCAGG CATGTGTCCC TTCCAAGCTG TGTTAACTGT TCAAACTCAG GCCTGTGTGA   2717

CTCCATTGGG GTGAGAGGTG AAAGCATAAC ATGGGTACAG AGGGGACAAC AATGAATCAG   2777

AACAGATGCT GAGCCATAGG TCTAAATAGG ATCCTGGAGG CTGCCTGCTG TGCTGGGAGG   2837

TATAGGGGTC CTGGGGGCAG GCCAGGGCAG TTGACAGGTA CTTGGAGGGC TCAGGGCAGT   2897

GGCTTCTTTC CAGTATGGAA GGATTTCAAC ATTTTAATAG TTGGTTAGGC TAAACTGGTG   2957

CATACTGGCA TTGGCCTTGG TGGGGAGCAC AGACACAGGA TAGGACTCCA TTTCTTTCTT   3017

CCATTCCTTC ATGTCTAGGA TAACTTGCTT TCTTCTTTCC TTTACTCCTG GCTCAAGCCC   3077

TGAATTTCTT CTTTTCCTGC AGGGGTTGAG AGCTTTCTGC CTTAGCCTAC CATGTGAAAC   3137

TCTACCCTGA AGAAAGGGAT GGATAGGAAG TAGACCTCTT TTTCTTACCA GTCTCCTCCC   3197

CTACTCTGCC CCCTAAGCTG GCTGTACCTG TTCCTCCCCC ATAAAATGAT CCTGCCAATC   3257

TAAAAAAAAA A                                                       3268
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 851 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Gln Trp Glu Met Leu Gln Asn Leu Asp Ser Pro Phe Gln Asp
 1               5                  10                  15

Gln Leu His Gln Leu Tyr Ser His Ser Leu Leu Pro Val Asp Ile Arg
            20                  25                  30

Gln Tyr Leu Ala Val Trp Ile Glu Asp Gln Asn Trp Gln Glu Ala Ala
        35                  40                  45

Leu Gly Ser Asp Asp Ser Lys Ala Thr Met Leu Phe Phe His Phe Leu
    50                  55                  60

Asp Gln Leu Asn Tyr Glu Cys Gly Arg Cys Ser Gln Asp Pro Glu Ser
65                  70                  75                  80

Leu Leu Leu Gln His Asn Leu Arg Lys Phe Cys Arg Asp Ile Gln Pro
                85                  90                  95

Phe Ser Gln Asp Pro Thr Gln Leu Ala Glu Met Ile Phe Asn Leu Leu
            100                 105                 110

Leu Glu Glu Lys Arg Ile Leu Ile Gln Ala Gln Arg Ala Gln Leu Glu
        115                 120                 125

Gln Gly Glu Pro Val Leu Glu Thr Pro Val Glu Ser Gln Gln His Glu
    130                 135                 140

Ile Glu Ser Arg Ile Leu Asp Leu Arg Ala Met Met Glu Lys Leu Val
145                 150                 155                 160

Lys Ser Ile Ser Gln Leu Lys Asp Gln Gln Asp Val Phe Cys Phe Arg
                165                 170                 175

Tyr Lys Ile Gln Ala Lys Gly Lys Thr Pro Ser Leu Asp Pro His Gln
            180                 185                 190

Thr Lys Glu Gln Lys Ile Leu Gln Glu Thr Leu Asn Glu Leu Asp Lys
        195                 200                 205

Arg Arg Lys Glu Val Leu Asp Ala Ser Lys Ala Leu Leu Gly Arg Leu
    210                 215                 220

Thr Thr Leu Ile Glu Leu Leu Leu Pro Lys Leu Glu Glu Trp Lys Ala
```

-continued

```
225                 230                 235                 240

Gln Gln Gln Lys Ala Cys Ile Arg Ala Pro Ile Asp His Gly Leu Glu
                245                 250                 255

Gln Leu Glu Thr Trp Phe Thr Ala Gly Ala Lys Leu Leu Phe His Leu
            260                 265                 270

Arg Gln Leu Leu Lys Glu Leu Lys Gly Leu Ser Cys Leu Val Ser Tyr
        275                 280                 285

Gln Asp Asp Pro Leu Thr Lys Gly Val Asp Leu Arg Asn Ala Gln Val
    290                 295                 300

Thr Glu Leu Leu Gln Arg Leu Leu His Arg Ala Phe Val Val Glu Thr
305                 310                 315                 320

Gln Pro Cys Met Pro Gln Thr Pro His Arg Pro Leu Ile Leu Lys Thr
                325                 330                 335

Gly Ser Lys Phe Thr Val Arg Thr Arg Leu Leu Val Arg Leu Gln Glu
            340                 345                 350

Gly Asn Glu Ser Leu Thr Val Glu Val Ser Ile Asp Arg Asn Pro Pro
        355                 360                 365

Gln Leu Gln Gly Phe Arg Lys Phe Asn Ile Leu Thr Ser Asn Gln Lys
    370                 375                 380

Thr Leu Thr Pro Glu Lys Gly Gln Ser Gln Gly Leu Ile Trp Asp Phe
385                 390                 395                 400

Gly Tyr Leu Thr Leu Val Glu Gln Arg Ser Gly Gly Ser Gly Lys Gly
                405                 410                 415

Ser Asn Lys Gly Pro Leu Gly Val Thr Glu Glu Leu His Ile Ile Ser
            420                 425                 430

Phe Thr Val Lys Tyr Thr Tyr Gln Gly Leu Lys Gln Glu Leu Lys Thr
        435                 440                 445

Asp Thr Leu Pro Val Val Ile Ile Ser Asn Met Asn Gln Leu Ser Ile
    450                 455                 460

Ala Trp Ala Ser Val Leu Trp Phe Asn Leu Leu Ser Pro Asn Leu Gln
465                 470                 475                 480

Asn Gln Gln Phe Phe Ser Asn Pro Pro Lys Ala Pro Trp Ser Leu Leu
                485                 490                 495

Gly Pro Ala Leu Ser Trp Gln Phe Ser Ser Tyr Val Gly Arg Gly Leu
            500                 505                 510

Asn Ser Asp Gln Leu Ser Met Leu Arg Asn Lys Leu Phe Gly Gln Asn
        515                 520                 525

Cys Arg Thr Glu Asp Pro Leu Leu Ser Trp Ala Asp Phe Thr Lys Arg
    530                 535                 540

Glu Ser Pro Pro Gly Lys Leu Pro Phe Trp Thr Trp Leu Asp Lys Ile
545                 550                 555                 560

Leu Glu Leu Val His Asp His Leu Lys Asp Leu Trp Asn Asp Gly Arg
                565                 570                 575

Ile Met Gly Phe Val Ser Arg Ser Gln Glu Arg Arg Leu Leu Lys Lys
            580                 585                 590

Thr Met Ser Gly Thr Phe Leu Leu Arg Phe Ser Glu Ser Ser Glu Gly
        595                 600                 605

Gly Ile Thr Cys Ser Trp Val Glu His Gln Asp Asp Asp Lys Val Leu
    610                 615                 620

Ile Tyr Ser Val Gln Pro Tyr Thr Lys Glu Val Leu Gln Ser Leu Pro
625                 630                 635                 640

Leu Thr Glu Ile Ile Arg His Tyr Gln Leu Leu Thr Glu Glu Asn Ile
                645                 650                 655
```

-continued

```
Pro Glu Asn Pro Leu Arg Phe Leu Tyr Pro Arg Ile Pro Arg Asp Glu
            660                 665                 670

Ala Phe Gly Cys Tyr Tyr Gln Glu Lys Val Asn Leu Gln Glu Arg Arg
        675                 680                 685

Lys Tyr Leu Lys His Arg Leu Ile Val Val Ser Asn Arg Gln Val Asp
    690                 695                 700

Glu Leu Gln Gln Pro Leu Glu Leu Lys Pro Glu Pro Glu Leu Glu Ser
705                 710                 715                 720

Leu Glu Leu Glu Leu Gly Leu Val Pro Glu Pro Glu Leu Ser Leu Asp
                725                 730                 735

Leu Glu Pro Leu Leu Lys Ala Gly Leu Asp Leu Gly Pro Glu Leu Glu
            740                 745                 750

Ser Val Leu Glu Ser Thr Leu Glu Pro Val Ile Glu Pro Thr Leu Cys
        755                 760                 765

Met Val Ser Gln Thr Val Pro Glu Pro Asp Gln Gly Pro Val Ser Gln
    770                 775                 780

Pro Val Pro Glu Pro Asp Leu Pro Cys Asp Leu Arg His Leu Asn Thr
785                 790                 795                 800

Glu Pro Met Glu Ile Phe Arg Asn Cys Val Lys Ile Glu Glu Ile Met
                805                 810                 815

Pro Asn Gly Asp Pro Leu Leu Ala Gly Gln Asn Thr Val Asp Glu Val
            820                 825                 830

Tyr Val Ser Arg Pro Ser His Phe Tyr Thr Asp Gly Pro Leu Met Pro
        835                 840                 845

Ser Asp Phe
    850
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3943 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: Human Stat91

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 197..2449

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATTAAACCTC TCGCCGAGCC CCTCCGCAGA CTCTGCGCCG GAAAGTTTCA TTTGCTGTAT     60

GCCATCCTCG AGAGCTGTCT AGGTTAACGT TCGCACTCTG TGTATATAAC CTCGACAGTC    120

TTGGCACCTA ACGTGCTGTG CGTAGCTGCT CCTTTGGTTG AATCCCCAGG CCCTTGTTGG    180

GGCACAAGGT GGCAGG ATG TCT CAG TGG TAC GAA CTT CAG CAG CTT GAC         229
                  Met Ser Gln Trp Tyr Glu Leu Gln Gln Leu Asp
                    1               5                  10

TCA AAA TTC CTG GAG CAG GTT CAC CAG CTT TAT GAT GAC AGT TTT CCC       277
Ser Lys Phe Leu Glu Gln Val His Gln Leu Tyr Asp Asp Ser Phe Pro
```

-continued

```
            15                  20                  25

ATG GAA ATC AGA CAG TAC CTG GCA CAG TGG TTA GAA AAG CAA GAC TGG      325
Met Glu Ile Arg Gln Tyr Leu Ala Gln Trp Leu Glu Lys Gln Asp Trp
         30                  35                  40

GAG CAC GCT GCC AAT GAT GTT TCA TTT GCC ACC ATC CGT TTT CAT GAC      373
Glu His Ala Ala Asn Asp Val Ser Phe Ala Thr Ile Arg Phe His Asp
     45                  50                  55

CTC CTG TCA CAG CTG GAT GAT CAA TAT AGT CGC TTT TCT TTG GAG AAT      421
Leu Leu Ser Gln Leu Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn
 60                  65                  70                  75

AAC TTC TTG CTA CAG CAT AAC ATA AGG AAA AGC AAG CGT AAT CTT CAG      469
Asn Phe Leu Leu Gln His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln
                 80                  85                  90

GAT AAT TTT CAG GAA GAC CCA ATC CAG ATG TCT ATG ATC ATT TAC AGC      517
Asp Asn Phe Gln Glu Asp Pro Ile Gln Met Ser Met Ile Ile Tyr Ser
             95                  100                 105

TGT CTG AAG GAA GAA AGG AAA ATT CTG GAA AAC GCC CAG AGA TTT AAT      565
Cys Leu Lys Glu Glu Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn
        110                 115                 120

CAG GCT CAG TCG GGG AAT ATT CAG AGC ACA GTG ATG TTA GAC AAA CAG      613
Gln Ala Gln Ser Gly Asn Ile Gln Ser Thr Val Met Leu Asp Lys Gln
    125                 130                 135

AAA GAG CTT GAC AGT AAA GTC AGA AAT GTG AAG GAC AAG GTT ATG TGT      661
Lys Glu Leu Asp Ser Lys Val Arg Asn Val Lys Asp Lys Val Met Cys
140                 145                 150                 155

ATA GAG CAT GAA ATC AAG AGC CTG GAA GAT TTA CAA GAT GAA TAT GAC      709
Ile Glu His Glu Ile Lys Ser Leu Glu Asp Leu Gln Asp Glu Tyr Asp
                160                 165                 170

TTC AAA TGC AAA ACC TTG CAG AAC AGA GAA CAC GAG ACC AAT GGT GTG      757
Phe Lys Cys Lys Thr Leu Gln Asn Arg Glu His Glu Thr Asn Gly Val
            175                 180                 185

GCA AAG AGT GAT CAG AAA CAA GAA CAG CTG TTA CTC AAG AAG ATG TAT      805
Ala Lys Ser Asp Gln Lys Gln Glu Gln Leu Leu Leu Lys Lys Met Tyr
        190                 195                 200

TTA ATG CTT GAC AAT AAG AGA AAG GAA GTA GTT CAC AAA ATA ATA GAG      853
Leu Met Leu Asp Asn Lys Arg Lys Glu Val Val His Lys Ile Ile Glu
    205                 210                 215

TTG CTG AAT GTC ACT GAA CTT ACC CAG AAT GCC CTG ATT AAT GAT GAA      901
Leu Leu Asn Val Thr Glu Leu Thr Gln Asn Ala Leu Ile Asn Asp Glu
220                 225                 230                 235

CTA GTG GAG TGG AAG CGG AGA CAG CAG AGC GCC TGT ATT GGG GGG CCG      949
Leu Val Glu Trp Lys Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro
                240                 245                 250

CCC AAT GCT TGC TTG GAT CAG CTG CAG AAC TGG TTC ACT ATA GTT GCG      997
Pro Asn Ala Cys Leu Asp Gln Leu Gln Asn Trp Phe Thr Ile Val Ala
            255                 260                 265

GAG AGT CTG CAG CAA GTT CGG CAG CAG CTT AAA AAG TTG GAG GAA TTG     1045
Glu Ser Leu Gln Gln Val Arg Gln Gln Leu Lys Lys Leu Glu Glu Leu
        270                 275                 280

GAA CAG AAA TAC ACC TAC GAA CAT GAC CCT ATC ACA AAA AAC AAA CAA     1093
Glu Gln Lys Tyr Thr Tyr Glu His Asp Pro Ile Thr Lys Asn Lys Gln
    285                 290                 295

GTG TTA TGG GAC CGC ACC TTC AGT CTT TTC CAG CAG CTC ATT CAG AGC     1141
Val Leu Trp Asp Arg Thr Phe Ser Leu Phe Gln Gln Leu Ile Gln Ser
300                 305                 310                 315

 TCG TTT GTG GTG GAA AGA CAG CCC TGC ATG CCA ACG CAC CCT CAG AGG    1189
Ser Phe Val Val Glu Arg Gln Pro Cys Met Pro Thr His Pro Gln Arg
                320                 325                 330

CCG CTG GTC TTG AAG ACA GGG GTC CAG TTC ACT GTG AAG TTG AGA CTG     1237
```

-continued

```
Pro Leu Val Leu Lys Thr Gly Val Gln Phe Thr Val Lys Leu Arg Leu
            335                 340                 345

TTG GTG AAA TTG CAA GAG CTG AAT TAT AAT TTG AAA GTC AAA GTC TTA     1285
Leu Val Lys Leu Gln Glu Leu Asn Tyr Asn Leu Lys Val Lys Val Leu
        350                 355                 360

TTT GAT AAA GAT GTG AAT GAG AGA AAT ACA GTA AAA GGA TTT AGG AAG     1333
Phe Asp Lys Asp Val Asn Glu Arg Asn Thr Val Lys Gly Phe Arg Lys
    365                 370                 375

TTC AAC ATT TTG GGC ACG CAC ACA AAA GTG ATG AAC ATG GAG GAG TCC     1381
Phe Asn Ile Leu Gly Thr His Thr Lys Val Met Asn Met Glu Glu Ser
380                 385                 390                 395

ACC AAT GGC AGT CTG GCG GCT GAA TTT CGG CAC CTG CAA TTG AAA GAA     1429
Thr Asn Gly Ser Leu Ala Ala Glu Phe Arg His Leu Gln Leu Lys Glu
                400                 405                 410

CAG AAA AAT GCT GGC ACC AGA ACG AAT GAG GGT CCT CTC ATC GTT ACT     1477
Gln Lys Asn Ala Gly Thr Arg Thr Asn Glu Gly Pro Leu Ile Val Thr
            415                 420                 425

GAA GAG CTT CAC TCC CTT AGT TTT GAA ACC CAA TTG TGC CAG CCT GGT     1525
Glu Glu Leu His Ser Leu Ser Phe Glu Thr Gln Leu Cys Gln Pro Gly
        430                 435                 440

TTG GTA ATT GAC CTC GAG ACG ACC TCT CTG CCC GTT GTG GTG ATC TCC     1573
Leu Val Ile Asp Leu Glu Thr Thr Ser Leu Pro Val Val Val Ile Ser
    445                 450                 455

AAC GTC AGC CAG CTC CCG AGC GGT TGG GCC TCC ATC CTT TGG TAC AAC     1621
Asn Val Ser Gln Leu Pro Ser Gly Trp Ala Ser Ile Leu Trp Tyr Asn
460                 465                 470                 475

ATG CTG GTG GCG GAA CCC AGG AAT CTG TCC TTC TTC CTG ACT CCA CCA     1669
Met Leu Val Ala Glu Pro Arg Asn Leu Ser Phe Phe Leu Thr Pro Pro
                480                 485                 490

TGT GCA CGA TGG GCT CAG CTT TCA GAA GTG CTG AGT TGG CAG TTT TCT     1717
Cys Ala Arg Trp Ala Gln Leu Ser Glu Val Leu Ser Trp Gln Phe Ser
            495                 500                 505

TCT GTC ACC AAA AGA GGT CTC AAT GTG GAC CAG CTG AAC ATG TTG GGA     1765
Ser Val Thr Lys Arg Gly Leu Asn Val Asp Gln Leu Asn Met Leu Gly
        510                 515                 520

GAG AAG CTT CTT GGT CCT AAC GCC AGC CCC GAT GGT CTC ATT CCG TGG     1813
Glu Lys Leu Leu Gly Pro Asn Ala Ser Pro Asp Gly Leu Ile Pro Trp
    525                 530                 535

ACG AGG TTT TGT AAG GAA AAT ATA AAT GAT AAA AAT TTT CCC TTC TGG     1861
Thr Arg Phe Cys Lys Glu Asn Ile Asn Asp Lys Asn Phe Pro Phe Trp
540                 545                 550                 555

CTT TGG ATT GAA AGC ATC CTA GAA CTC ATT AAA AAA CAC CTG CTC CCT     1909
Leu Trp Ile Glu Ser Ile Leu Glu Leu Ile Lys Lys His Leu Leu Pro
                560                 565                 570

CTC TGG AAT GAT GGG TGC ATC ATG GGC TTC ATC AGC AAG GAG CGA GAG     1957
Leu Trp Asn Asp Gly Cys Ile Met Gly Phe Ile Ser Lys Glu Arg Glu
            575                 580                 585

 CGT GCC CTG TTG AAG GAC CAG CAG CCG GGG ACC TTC CTG CTG CGG TTC    2005
Arg Ala Leu Leu Lys Asp Gln Gln Pro Gly Thr Phe Leu Leu Arg Phe
        590                 595                 600

AGT GAG AGC TCC CGG GAA GGG GCC ATC ACA TTC ACA TGG GTG GAG CGG     2053
Ser Glu Ser Ser Arg Glu Gly Ala Ile Thr Phe Thr Trp Val Glu Arg
    605                 610                 615

TCC CAG AAC GGA GGC GAA CCT GAC TTC CAT GCG GTT GAA CCC TAC ACG     2101
Ser Gln Asn Gly Gly Glu Pro Asp Phe His Ala Val Glu Pro Tyr Thr
620                 625                 630                 635

AAG AAA GAA CTT TCT GCT GTT ACT TTC CCT GAC ATC ATT CGC AAT TAC     2149
Lys Lys Glu Leu Ser Ala Val Thr Phe Pro Asp Ile Ile Arg Asn Tyr
                640                 645                 650
```

-continued

```
AAA GTC ATG GCT GCT GAG AAT ATT CCT GAG AAT CCC CTG AAG TAT CTG     2197
Lys Val Met Ala Ala Glu Asn Ile Pro Glu Asn Pro Leu Lys Tyr Leu
            655                 660                 665

TAT CCA AAT ATT GAC AAA GAC CAT GCC TTT GGA AAG TAT TAC TCC AGG     2245
Tyr Pro Asn Ile Asp Lys Asp His Ala Phe Gly Lys Tyr Tyr Ser Arg
        670                 675                 680

CCA AAG GAA GCA CCA GAG CCA ATG GAA CTT GAT GGC CCT AAA GGA ACT     2293
Pro Lys Glu Ala Pro Glu Pro Met Glu Leu Asp Gly Pro Lys Gly Thr
    685                 690                 695

GGA TAT ATC AAG ACT GAG TTG ATT TCT GTG TCT GAA GTT CAC CCT TCT     2341
Gly Tyr Ile Lys Thr Glu Leu Ile Ser Val Ser Glu Val His Pro Ser
700                 705                 710                 715

AGA CTT CAG ACC ACA GAC AAC CTG CTC CCC ATG TCT CCT GAG GAG TTT     2389
Arg Leu Gln Thr Thr Asp Asn Leu Leu Pro Met Ser Pro Glu Glu Phe
                720                 725                 730

GAC GAG GTG TCT CGG ATA GTG GGC TCT GTA GAA TTC GAC AGT ATG ATG     2437
Asp Glu Val Ser Arg Ile Val Gly Ser Val Glu Phe Asp Ser Met Met
            735                 740                 745

AAC ACA GTA TAGAGCATGA ATTTTTTTCA TCTTCTCTGG CGACAGTTTT             2486
Asn Thr Val
        750

CCTTCTCATC TGTGATTCCC TCCTGCTACT CTGTTCCTTC ACATCCTGTG TTTCTAGGGA   2546

AATGAAAGAA AGGCCAGCAA ATTCGCTGCA ACCTGTTGAT AGCAAGTGAA TTTTTCTCTA   2606

ACTCAGAAAC ATCAGTTACT CTGAAGGGCA TCATGCATCT TACTGAAGGT AAAATTGAAA   2666

GGCATTCTCT GAAGAGTGGG TTTCACAAGT GAAAAACATC CAGATACACC CAAAGTATCA   2726

GGACGAGAAT GAGGGTCCTT TGGGAAAGGA GAAGTTAAGC AACATCTAGC AAATGTTATG   2786

CATAAAGTCA GTGCCCAACT GTTATAGGTT GTTGGATAAA TCAGTGGTTA TTTAGGGAAC   2846

TGCTTGACGT AGGAACGGTA AATTTCTGTG GGAGAATTCT TACATGTTTT CTTTGCTTTA   2906

AGTGTAACTG GCAGTTTTCC ATTGGTTTAC CTGTGAAATA GTTCAAAGCC AAGTTTATAT   2966

ACAATTATAT CAGTCCTCTT TCAAAGGTAG CCATCATGGA TCTGGTAGGG GGAAAATGTG   3026

TATTTTATTA CATCTTTCAC ATTGGCTATT TAAAGACAAA GACAAATTCT GTTTCTTGAG   3086

AAGAGAACAT TTCCAAATTC ACAAGTTGTG TTTGATATCC AAAGCTGAAT ACATTCTGCT   3146

TTCATCTTGG TCACATACAA TTATTTTTAC AGTTCTCCCA AGGGAGTTAG GCTATTCACA   3206

ACCACTCATT CAAAAGTTGA AATTAACCAT AGATGTAGAT AAACTCAGAA ATTTAATTCA   3266

TGTTTCTTAA ATGGGCTACT TTGTCCTTTT TGTTATTAGG GTGGTATTTA GTCTATTAGC   3326

CACAAAATTG GGAAAGGAGT AGAAAAAGCA GTAACTGACA ACTTGAATAA TACACCAGAG   3386

ATAATATGAG AATCAGATCA TTTCAAAACT CATTTCCTAT GTAACTGCAT TGAGAACTGC   3446

ATATGTTTCG CTGATATATG TGTTTTTCAC ATTTGCGAAT GGTTCCATTC TCTCTCCTGT   3506

ACTTTTTCCA GACACTTTTT TGAGTGGATG ATGTTTCGTG AAGTATACTG TATTTTTACC   3566

TTTTTCCTTC CTTATCACTG ACACAAAAAG TAGATTAAGA GATGGGTTTG ACAAGGTTCT   3626

TCCCTTTTAC ATACTGCTGT CTATGTGGCT GTATCTTGTT TTTCCACTAC TGCTACCACA   3686

ACTATATTAT CATGCAAATG CTGTATTCTT CTTTGGTGGA GATAAAGATT TCTTGAGTTT   3746

TGTTTTAAAA TTAAAGCTAA AGTATCTGTA TTGCATTAAA TATAATATCG ACACAGTGCT   3806

TTCCGTGGCA CTGCATACAA TCTGAGGCCT CCTCTCTCAG TTTTTATATA GATGGCGAGA   3866

ACCTAAGTTT CAGTTGATTT TACAATTGAA ATGACTAAAA AACAAAGAAG ACAACATTAA   3926

AAACAATATT GTTTCTA                                                  3943
```

-continued

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 750 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ser Gln Trp Tyr Glu Leu Gln Gln Leu Asp Ser Lys Phe Leu Glu
 1               5                  10                  15

Gln Val His Gln Leu Tyr Asp Asp Ser Phe Pro Met Glu Ile Arg Gln
            20                  25                  30

Tyr Leu Ala Gln Trp Leu Glu Lys Gln Asp Trp Glu His Ala Ala Asn
        35                  40                  45

Asp Val Ser Phe Ala Thr Ile Arg Phe His Asp Leu Leu Ser Gln Leu
    50                  55                  60

Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn Asn Phe Leu Leu Gln
65                  70                  75                  80

His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln Asp Asn Phe Gln Glu
                85                  90                  95

Asp Pro Ile Gln Met Ser Met Ile Ile Tyr Ser Cys Leu Lys Glu Glu
            100                 105                 110

Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn Gln Ala Gln Ser Gly
        115                 120                 125

Asn Ile Gln Ser Thr Val Met Leu Asp Lys Gln Lys Glu Leu Asp Ser
    130                 135                 140

Lys Val Arg Asn Val Lys Asp Lys Val Met Cys Ile Glu His Glu Ile
145                 150                 155                 160

Lys Ser Leu Glu Asp Leu Gln Asp Glu Tyr Asp Phe Lys Cys Lys Thr
                165                 170                 175

Leu Gln Asn Arg Glu His Glu Thr Asn Gly Val Ala Lys Ser Asp Gln
            180                 185                 190

Lys Gln Glu Gln Leu Leu Leu Lys Lys Met Tyr Leu Met Leu Asp Asn
        195                 200                 205

Lys Arg Lys Glu Val Val His Lys Ile Ile Glu Leu Leu Asn Val Thr
    210                 215                 220

Glu Leu Thr Gln Asn Ala Leu Ile Asn Asp Glu Leu Val Glu Trp Lys
225                 230                 235                 240

Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro Pro Asn Ala Cys Leu
                245                 250                 255

Asp Gln Leu Gln Asn Trp Phe Thr Ile Val Ala Glu Ser Leu Gln Gln
            260                 265                 270

Val Arg Gln Gln Leu Lys Lys Leu Glu Glu Leu Glu Gln Lys Tyr Thr
        275                 280                 285

Tyr Glu His Asp Pro Ile Thr Lys Asn Lys Gln Val Leu Trp Asp Arg
    290                 295                 300

Thr Phe Ser Leu Phe Gln Gln Leu Ile Gln Ser Ser Phe Val Val Glu
305                 310                 315                 320

Arg Gln Pro Cys Met Pro Thr His Pro Gln Arg Pro Leu Val Leu Lys
                325                 330                 335

Thr Gly Val Gln Phe Thr Val Lys Leu Arg Leu Leu Val Lys Leu Gln
            340                 345                 350

Glu Leu Asn Tyr Asn Leu Lys Val Lys Val Leu Phe Asp Lys Asp Val
        355                 360                 365
```